Title

Structural and biochemical basis for regiospecificity of the flavonoid glycosyltransferase UGT95A1

Journal

Journal of Biological Chemistry, 300(9)

ISSN

0021-9258

Authors

Sirirungruang, Sasilada

Blay, Vincent

Scott, Yasmine F

et al.

Publication Date

2024-09-01

DOI

10.1016/j.jbc.2024.107602

Peer reviewed

JBC RESEARCH ARTICLE

Structural and biochemical basis for regiospecificity of the flavonoid glycosyltransferase UGT95A1

Received for publication, April 30, 2024, and in revised form, July 2, 2024 Published, Papers in Press, July 24, 2024,
https://doi.org/10.1016/j.jbc.2024.107602

Sasilada Sirirungruang[1,2,3,4], Vincent Blay[5], Yasmine F. Scott[6], Jose H. Pereira[7,8], Michal Hammel[8], Collin R. Barnum[9], Paul D. Adams[7,8], and Patrick M. Shih[1,2,3,10,*]

From the [1]Department of Plant and Microbial Biology, University of California, Berkeley, California, USA; [2]Feedstocks Division, Joint BioEnergy Institute, Emeryville, California, USA; [3]Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, California, USA; [4]Center for Biomolecular Structure, Function and Application, Suranaree University of Technology, Nakhon Ratchasima, Thailand; [5]Biofuels and Bioproducts Division, Joint BioEnergy Institute, Emeryville, California, USA; [6]Department of Molecular and Cell Biology, University of California, Berkeley, California, USA; [7]Technology Division, Joint BioEnergy Institute, Emeryville, California, USA; [8]Molecular Biophysics and Integrated Bioimaging Division, Lawrence Berkeley National Laboratory, Berkeley, California, USA; [9]Biochemistry, Molecular, Cellular and Developmental Biology Graduate Group, University of California, Davis, California, USA; [10]Innovative Genomics Institute, University of California, Berkeley, California, USA

Reviewed by members of the JBC Editorial Board. Edited by Sarah E. O'Connor

Glycosylation is a predominant strategy plants use to fine-tune the properties of small molecule metabolites to affect their bioactivity, transport, and storage. It is also important in biotechnology and medicine as many glycosides are utilized in human health. Small molecule glycosylation is largely carried out by family 1 glycosyltransferases. Here, we report a structural and biochemical investigation of UGT95A1, a family 1 GT enzyme from *Pilosella officinarum* that exhibits a strong, unusual regiospecificity for the 3′-*O* position of flavonoid acceptor substrate luteolin. We obtained an apo crystal structure to help drive the analyses of a series of binding site mutants, revealing that while most residues are tolerant to mutations, key residues M145 and D464 are important for overall glycosylation activity. Interestingly, E347 is crucial for maintaining the strong preference for 3′-*O* glycosylation, while R462 can be mutated to increase regioselectivity. The structural determinants of regioselectivity were further confirmed in homologous enzymes. Our study also suggests that the enzyme contains large, highly dynamic, disordered regions. We showed that while most disordered regions of the protein have little to no implication in catalysis, the disordered regions conserved among investigated homologs are important to both the overall efficiency and regiospecificity of the enzyme. This report represents a comprehensive in-depth analysis of a family 1 GT enzyme with a unique substrate regiospecificity and may provide a basis for enzyme functional prediction and engineering.

Glycosylation is a unique facet of plant metabolism. Small molecule glycosylation is arguably the predominant strategy plants use to expand their metabolome chemical diversity (1). Glycosylation can modify a molecule's hydrophobicity, stability, and bioactivity, which in turn affects its subcellular localization, transport, and storage (2–4). Thus, the strategy is widely used by plants to control metabolite function and utility. Indeed, small molecule glycosylation has been shown to be important for growth and development (5, 6), communication (4), and defense (7, 8). As such, understanding glycosylation can offer important insights into plant physiology.

Small molecule glycosylation is also important in biotechnology (9). Many glycoside molecules are high-value chemicals widely utilized in human health due to their unique bioactivities. For example, cardiac glycosides are an important treatment for cardiac arrhythmia (10); steviol glycosides are widely used, noncaloric sweeteners (11, 12); and saponins, glycosides of triterpenes and sterols, are heavily studied and utilized for their anticancer, antiinflammatory, and immunomodulatory effects (13, 14). As their complex carbohydrate modules complicate the application of traditional synthetic methods, many of these molecules are extracted directly from the native producers, which may not be readily available. Therefore, there is much interest in elucidating the biosynthetic processes of this class of molecules (15).

The glycosylation steps in plant glycoside biosynthesis are mostly mediated by family 1 glycosyltransferases (GTs) according to CAZy classification (www.cazy.org) (16). Family 1 GT enzymes are inverting Leloir-type glycosyltransferases that glycosylate small molecules using uridine diphosphate (UDP) sugars as the sugar donor. They mostly use a histidine-aspartate catalytic dyad to deprotonate the nucleophile on the sugar acceptor, which then performs an S_N2-like attack on the anomeric carbon of the sugar donor (17). Family 1 GT enzymes also contain a conserved 44-amino acid-long motif called the "plant secondary product glycosylation" (PSPG) box, which binds to the nucleotide sugar donor substrate (17). Family 1 GTs have been shown to facilitate the formation of *O*-, *N*-, *S*-, and *C*-glycosides of a large repertoire of sugar acceptor substrates (18–20).

Family 1 GT enzymes are highly expanded in the plant kingdom, with many higher plants encoding over a hundred homologs in their genomes (21–24). While thousands of

* For correspondence: Patrick M. Shih, pmshih@berkeley.edu.

family 1 GT enzymes have been sequenced, only a small fraction of them have been studied in detail. The structure-function relationship controlling the sugar acceptor substrates and the enzyme regiospecificity of family 1 GTs is not well understood. In this work, we studied the enzyme UGT95A1 from *Pilosella officinarum* or mouse-ear hawkweed (previously *Hieracium pilosella* L.). This enzyme has previously been reported to transfer the glucose moiety from UDP-glucose to the flavone luteolin (25). Unlike most characterized flavonoid GT enzymes that glycosylate at the 3-*O* or the 7-*O* position, UGT95A1 was shown to have a strong specificity for the 3′-*O* position. UGT95A1 belongs to the clade Q of plant UGT enzymes, which have not previously been structurally characterized (25–27). We utilized a range of structural and biochemical approaches to elucidate the mechanism underlying UGT95A1 efficiency and regioselectivity. Using a combination of crystallography, mutational analysis, and kinetic study, we uncovered structural features and proposed the mechanism responsible for the behavior of this enzyme.

Results

UGT95A1 mediates single and double glycosylation of luteolin with a strong regioselectivity for the 3′-O position

To investigate the activity of UGT95A1, the enzyme was expressed heterologously in *Escherichia coli* and purified by affinity chromatography to near homogeneity (Fig. S1). The enzyme was incubated with luteolin, which can be glycosylated in three different positions (Fig. 1*A*), and UDP-glucose at 25 °C for 1 h. The resulting reaction products were analyzed using LC-MS. We observed mass features corresponding to three single glycosylation products: one major product, luteolin-3′-*O*-glucoside, and two minor products, luteolin-4′-*O*-glucoside and luteolin-7-*O*-glucoside (Fig. 1*B*). In addition, one double glycosylation product, luteolin-7,3′-di-*O*-glucoside, was observed (Fig. 1*B*). Our result is consistent with a previous report that showed that UGT95A1 produced luteolin-3′-*O*-glucoside (25). The minor products observed here are likely due to the use of a higher concentration of the enzyme in this study.

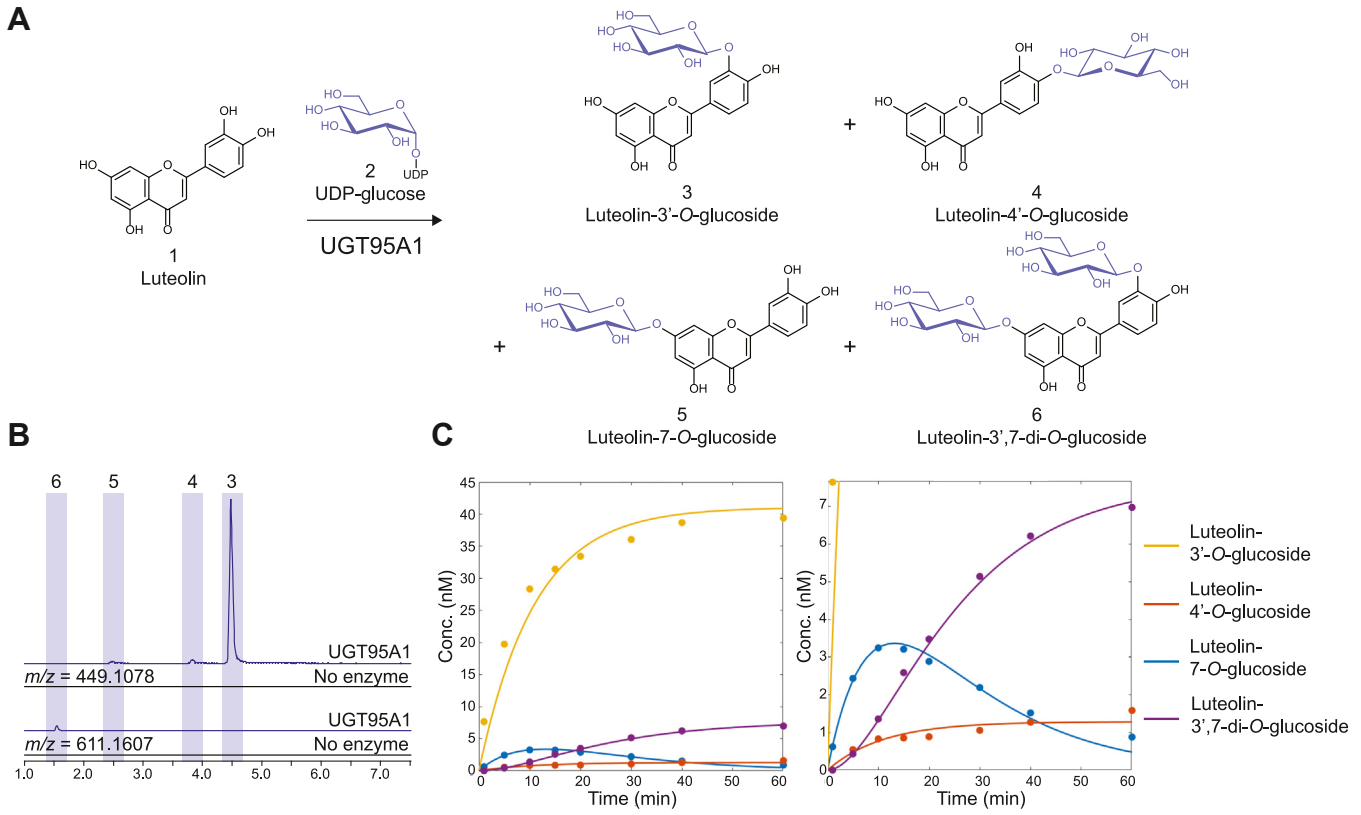

Figure 1. Luteolin glycosylation activity of UGT95A1. *A*, UGT95A1 facilitates the transfer of glucose from UDP-glucose sugar donor to nucleophilic hydroxyl groups of sugar acceptor luteolin. The reaction produces three single glycosylation products identified as luteolin-7-*O*-glucoside, luteolin-4′-*O*-glucoside, and luteolin-3′-*O*-glucoside, and one double glycosylation product, luteolin-3′,7-di-glucoside. *B*, LC-MS chromatograms of luteolin glycosylation reaction mediated by UGT95A1 show three chromatographically resolved single glycosylation products at *m/z* = 449.1078. They were identified as luteolin-7-*O*-glucoside at 2.5 min, luteolin-4′-*O*-glucoside at 3.9 min, and luteolin-3′-*O*-glucoside at 4.5 min, which is the dominant reaction product. One double glycosylation product was observed at *m/z* = 611.1607 and identified as luteolin-3′,7-di-glucoside. Chromatograms are representative of at least three technical replicates. *C*, kinetic analysis of luteolin glycosylation reaction mediated by UGT95A1 shows the amount of each product observed over time. UGT95A1 produces luteolin-3′-*O*-glucoside at the fastest rate. Interestingly, the amount of luteolin-7-*O*-glucoside decreases after an initial accumulation. This decrease corresponds with the appearance of luteolin-3′,7-di-glucoside. Reactions consisting of 250 nM enzyme, 50 μM luteolin, and 1.6 mM UDP-glucose were incubated at 25 °C for 1, 5, 10, 15, 20, 30, 40, and 60 min, at which time aliquots of 10 μl of the reactions were quenched with 20 μl of ice-cold 60 mM trichloroacetic acid. Quenched reactions were analyzed by LC-MS. *Solid circles* represent observed concentrations. *Solid lines* represent fitted kinetic models.

Kinetic analysis shows that UGT95A1 produces the major product luteolin-3′-*O*-glucoside at a rate roughly 5 and 30 times faster than the minor products luteolin-7-*O*-glucoside and luteolin-4′-*O*-glucoside, respectively (Figs. 1*C*, S2, and Table S2). Instead of the widely used continuous spectrometric assays that quantify the overall glycosylation rate, we performed a discontinuous LC-MS-based assay, which allows the assessment of the rate of each product formation individually. To calculate the rate of each product formation, reaction mixtures were sampled at designated time points, and products were separated and quantified by LC-MS. Besides the observed reaction rates, we found that the concentration of luteolin-7-*O*-glucoside initially increases and then decreases over time. This decrease correlates with the rise of the double glycosylation product, suggesting that luteolin-7-*O*-glucoside is further glycosylated to luteolin-7,3′-di-*O*-glucoside. With this assumption, the specific rate of the diglucoside formation is estimated to be 50 to 80% of that of luteolin-3′-*O*-glucoside formation (Table S2). Interestingly, the rate of diglucoside formation, which is a 3′-*O* glycosylation reaction of luteolin-7-*O*-glucoside, is faster than the rates of 7-*O* and 4′-*O* glycosylation of the preferred substrate luteolin. This observation suggests that UGT95A1 asserts a stronger regiospecificity than substrate selectivity, which is consistent with previous findings regarding the selectivity of related family 1 GT enzymes (17).

Structural analysis of UGT95A1 shows the expected overall GT-B fold and extensive disordered regions

To better understand the structural factors that govern the enzymatic behavior, we solved the apo crystal structure of UGT95A1. The three-dimensional structure was determined using molecular replacement and refined to 2.36 Å resolution (Protein Data Bank (PDB) ID: 9BCM). X-ray data collection and refinement statistics are presented in Table S3.

The structure of UGT95A1 consists of two N- and C-terminal Rossman-type domains and belongs to the GT-B fold (Fig. 2, *A* and *B*). The observed overall fold is similar to those of previously reported plant family 1 GT enzymes even though UGT95A1 exhibits low sequence similarity to them. The highest percent identity between UGT95A1 and a reported structure with high sequence alignment coverage (>80%) is only 28.23% to cyanohydrin *beta*-glucosyltransferase from *Sorghum bicolor* (PDB ID: 7ZER) (28). The N-terminal domain (residues 1–302) contains a central 7-stranded parallel β sheet decorated by 7 α-helices on both sides. The C-terminal

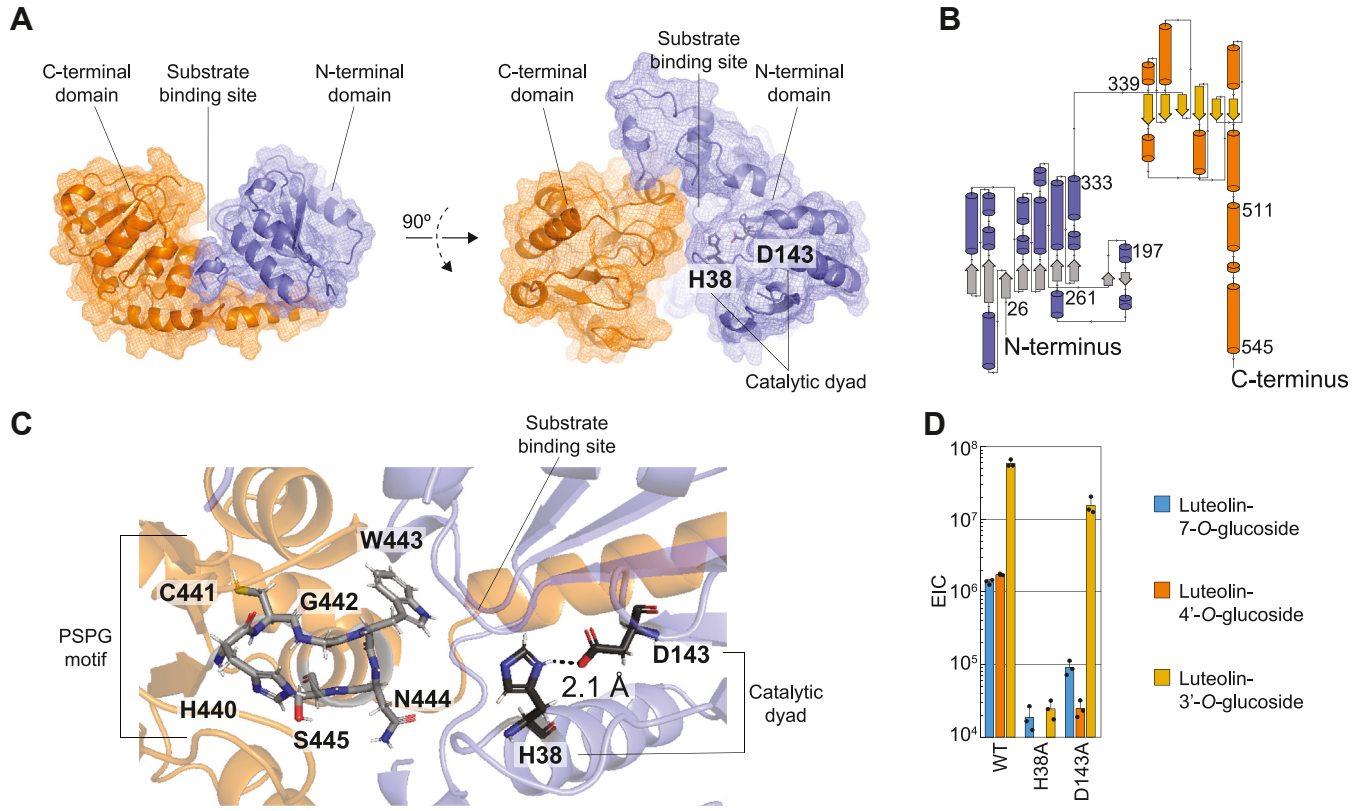

Figure 2. Structural determination of UGT95A1. *A*, crystal structure of UGT95A1 shows the expected GT-B overall fold, consisting of two Rossman-like domains in the N terminus (*blue*) and the C terminus (*orange*). The expected catalytic dyad H38 and D143 (shown in *sticks*) are found next to the expected substrate binding site between the N-terminal and the C-terminal domain. *B*, topological diagram of UGT95A1 based on crystal structure shows two domains (N-terminal domain in *blue*; C-terminal domain in *orange*), each of which consists of a large beta sheet decorated by helices. Amino acid residues at the beginning and the end of each domain are marked by residue numbers. *C*, the active site of UGT95A1 is made up of the catalytic dyad in the N-terminal domain (shown in *black sticks*) and the central part of the PSPG box, the HCGWNS motif, in the C-terminal domain (shown in *gray sticks*). *D*, extracted ion chromatogram (EIC) integration of luteolin glycosylation products produced by mutants of UGT95A1 lacking H38 or D143 of the catalytic dyad shows much decreased catalytic activity compared to the WT enzyme. Data are mean ± s.d. of three technical replicates. PSPG, plant secondary product glycosylation.

domain (residues 339–545) contains a 6-stranded β sheet surrounded by 7 α-helices. Extensive interactions between the N- and the C-terminal domains are observed including those between S166/G167-W443, E199-H475, A201-K470, R209-F466/D467, L274-I450, P295-A447, L297-E448, and F301-R452. The two domains pack tightly to form a cleft in the middle where the UDP-sugar and the acceptor substrate are expected to bind (Fig. 2*A*). The PSPG signature motif consisting of residues from W422 to Q465 lies within the C-terminal domain and is observed near the expected substrate binding cleft (Fig. 2*C*). H38 and D143 are found near the substrate binding pocket (Fig. 2*C*) and are expected to be the catalytic dyad essential for UGT95A1's glycosylation activity. Mutation of H38 to alanine virtually abolished the enzyme activity (Fig. 2*D*). Replacing D143 with alanine also severely compromised enzyme activity (Fig. 2*D*). These results confirm the role of the dyad in catalysis.

UGT95A1 likely contains multiple disordered regions (DRs). The apo structure lacks electron density for six regions of amino acids, namely residues 1 to 22 (DR1), 89 to 116 (DR2), 211 to 246 (DR3), 303 to 317 (DR4), 376 to 402 (DR5), and 484 to 494 (DR6), suggesting that up to 25% of the protein may be structurally disordered, flexible, or variable. Attempts to obtain substrate-bound structures to determine whether these regions become structured upon substrate binding were not successful, likely due to low affinity and solubility limit. We thus tried to determine the conformation of the DRs using computational protein structural prediction and molecular dynamics. AlphaFold2 (https://alphafold.ebi.ac.uk/) (29) predicted the coordinates of the majority of residues resolved in the crystal structure with high confidence; however, the residues that were not resolved in the crystal structure were predicted mostly with low confidence (pLDDT < 50). The top five predictions (Fig. S3) exhibited the same overall fold observed in the crystal structure and were generally very similar to one another (RMSD of all Cα of 2.0–3.9 Å; RMSD of high confidence Cα of 0.2–0.4 Å). The complete structural models predicted by AlphaFold2 were used as starting points for molecular dynamics simulations using Amber22 (30) and the ff19SB Amber force field (31). Molecular dynamic simulations totaling 3000 ns showed DR1, 2, and 3 to undergo rapid movements relative to the structured protein core (File S1). Interestingly, the simulations showed DR2, 3, and 5 to be interacting with one another and with the structured part of the protein.

To verify the computational predictions, we carried out size-exclusion chromatography coupled with small-angle X-ray scattering and multiangle light scattering (SEC-SAXS-MALS) experiments (32) on UGT95A1. SEC-SAXS-MALS analysis of UGT95A1 revealed an average experimental molecular weight of 52.27 kDa ± 0.063%, similar to the theoretical molecular weight of 59.3 kDa, suggesting the enzyme is monomeric in solution. (Simple Scattering Code: XSEXBUSH). The determined Guinier radius of gyration was 26.58 ± 0.82 Å. To determine the solution conformations of UGT95A1, each of the three 1000 ns trajectories of the enzyme was clustered by MDTraj into 30 clusters, and all 90 resulting clusters were fit to the SAXS dataset using FoXS (33, 34). The top three clusters whose theoretical scattering curves most closely match the experimental data set (χ^2 = 1.22, 1.27, and 1.34) have theoretical radii of gyration of 25.22, 24.89, and 25.14 Å, respectively, and all show the DRs to interact with one another and with the structured protein core *via* side-chain charge-charge interactions (Fig. S4). The interactions seem to keep the DRs in partially extended conformations, rather than fully extended into the bulk solvent. Collectively, these results suggest that UGT95A1 comprises a compact structured core with the highly conserved GT-B fold as well as putative disordered loop regions in partially extended conformations.

Active site residues play key roles in enzyme efficiency and regioselectivity

To determine residues important for enzymatic activity and regioselectivity, residues surrounding the expected substrate binding site were selected and mutated to alanine (Fig. 3*A*). The acceptor binding cavity is mainly made up of polar residues, unlike many other GT enzymes in which the pockets are made up of mostly hydrophobic residues (35, 36). Seventeen mutants were cloned, expressed, and purified to near homogeneity (Fig. S1). Most of the alanine mutants tested exhibited very similar behavior to the parental enzyme; however, four displayed altered activity profiles (Fig. 3*B*). Variants M145A and D464A had drastically affected enzymatic activity resulting in the reduced formation of all products. Variant E347A appeared to switch specificity from 3′-*O* glycosylation to 7-*O* glycosylation. On the other hand, R462A was observed to be more specific for 3′-*O* glycosylation, as shown in the reduced amounts of the minor products compared to the WT enzyme.

To further explore the role of these residues, we built small substitution libraries by altering the characteristics of the amino acids at the four selected positions. We found that, although substituting M145 with an alanine severely affected enzyme activity, other hydrophobic residues were well tolerated at the position (Fig. 3*C*). As M145 is found in close proximity (<4 Å) to the catalytic dyad (H38 and D143), we reason that it may help position the catalytic dyad for optimal acceptor substrate deprotonation.

Moreover, we found that when D464 was substituted by other amino acids of hydrophobic or positively charged nature, the mutant enzymes lost at least 95% of activity compared to WT (Fig. 3*D*). The only mutant that maintained a detectable level of activity was D464E, which maintains a negative charge upon substitution. Thus, we reasoned that a negative charge at position 464 is crucial for enzymatic activity. Given its position in the PSPG motif, we hypothesized that D464 may have a role in positioning the UDP-glucose for glycosylation. A visual investigation of the substrate bound structure of a related family 1 GT enzyme *Vv*GT1 (PDB ID: 2Z1C) (37) showed the corresponding residue (D374) coordinating the sugar moiety of the bound UDP-2-deoxy-2-fluoro glucose substrate analog. Taken together, these observations suggest that the D464

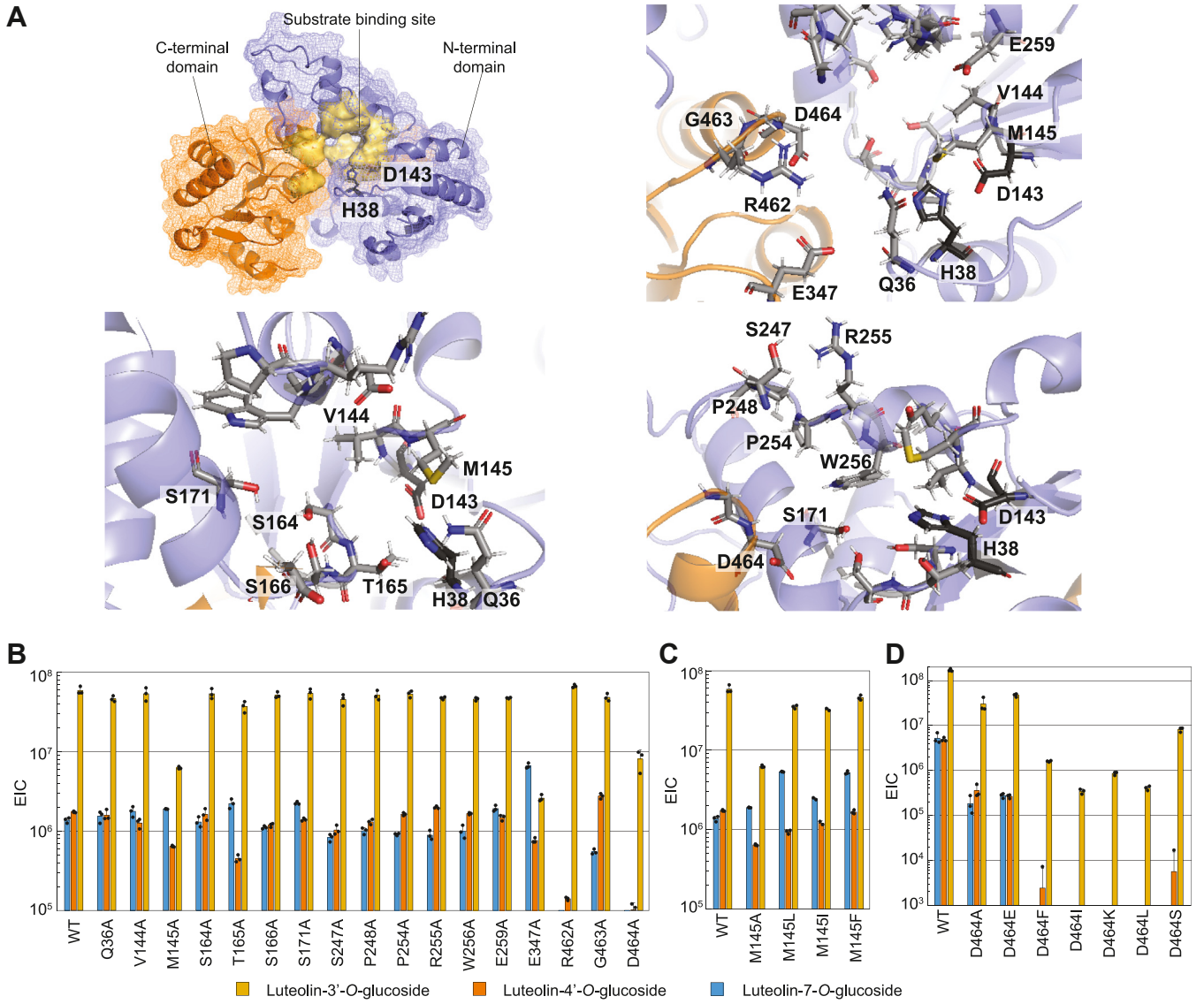

Figure 3. Contributions of active site residues to UGT95A1 enzymatic activity and regioselectivity. *A*, residues surrounding the expected acceptor substrate binding site (shown in *yellow* surface in *top left* panel and *gray sticks* in other panels) were chosen and mutated to alanine. *B*, EIC integration of luteolin glycosylation products produced by alanine mutants of substrate binding site residues of UGT95A1 shows that while most mutants exhibit WT-like behaviors, alanine mutants of M145, E347, R462, and D464 residues have activity profile drastically different from WT. Data are mean ± s.d. of three technical replicates. *C* and *D*, EIC integration of luteolin glycosylation products produced by M145 (*C*) and D464 (*D*) mutant libraries of UGT95A1 suggest that both residues are necessary for overall enzymatic efficiency. While M145 is tolerant of substitutions by other large hydrophobic residues, substitutions to D464 by any other amino acids drastically reduce enzymatic activity. Data are mean ± s.d. of three technical replicates. EIC, extracted ion chromatogram.

residue of UGT95A1 is important for positioning the glucose in the glycosylation reaction.

Unlike substitutions at M145 and D464, which negatively affected overall enzyme activity, mutations at position E347 appeared to shift the enzyme selectivity and favor the generation of luteolin-7-*O*-glucoside. From the E347 substitution library, we found that replacing the glutamate with neutral, hydrophobic residues such as alanine, leucine, isoleucine, and phenylalanine promoted glycosylation at the 7-*O* position of luteolin, whereas replacing it with aspartate did not result in the switch in regiospecificity (Fig. 4*A*). To further investigate the cause of regiospecificity change, we performed a kinetics study of the most active mutant, E347L. Examination of the rate constants revealed that E347L achieved its opposite regiospecificity by decreasing the rate of luteolin-3′-*O*-glucoside formation by roughly 100-fold while only decreasing the rate of luteolin-7-*O*-glucoside by half compared to the WT UGT95A1 (Figs. 4*B*, S5, and Table S2). This result suggests that E347 plays an important role in positioning luteolin substrate for nucleophilic attack leading to 3′-*O* glycosylation.

In contrast to E347, alanine mutant of R462 was found to promote 3′-*O* glycosylation and inhibit 7-*O* and 4′-*O* glycosylation, leading to increased regiospecificity compared to WT. Replacing R462 with other larger neutral, positively charged, or negatively charged residues also led to similar results observed for R462A (Fig. 4*C*). Kinetic study of R462A showed that the enzyme produced both minor products at rates 10 to 20 times

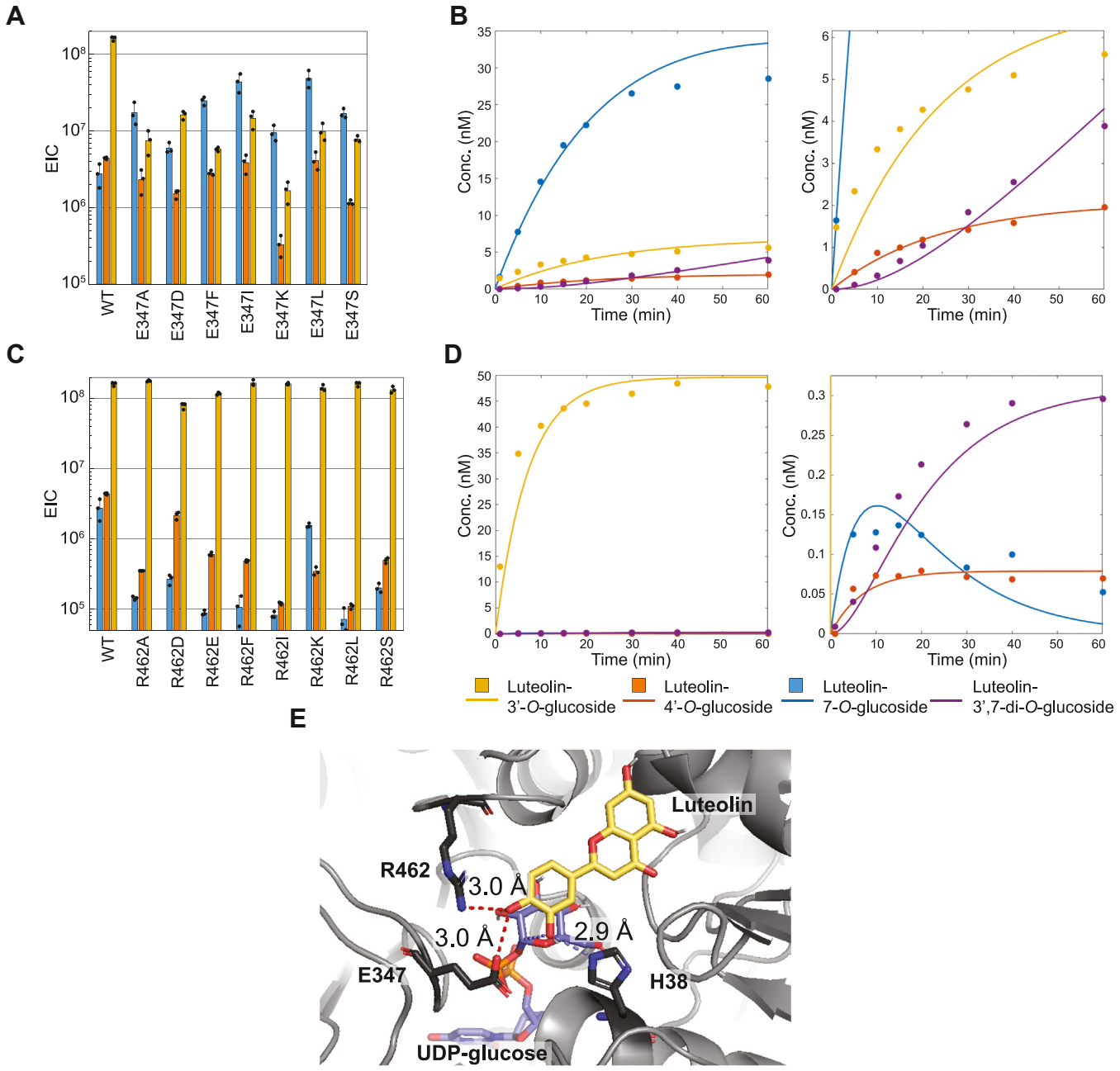

Figure 4. Roles of E347 and R462 residues in acceptor substrate regioselectivity. *A*, EIC integration of luteolin glycosylation products produced by E347 mutant libraries of UGT95A1 shows the residue to be important for maintaining a strong preference toward 3′-*O* glycosylation. Replacing E347 by non-negatively charged residues led to a change in regioselectivity to preferring 7-*O* glycosylation. Data are mean ± s.d. of three technical replicates. *B*, kinetic analysis of luteolin glycosylation reaction facilitated by UGT95A1 E347L shows the amount of each product observed over time. UGT95A1 E347L produces luteolin-7-*O*-glucoside at the fastest rate, while the production of luteolin-3′-*O*-glucoside is suppressed, resulting in a switch in regioselectivity from the WT enzyme. Reactions consisting of 1500 nM enzyme, 50 μM luteolin, and 1.6 mM UDP-glucose were incubated at 25 °C, quenched with ice-cold trichloroacetic acid, and analyzed by LC-MS. *Solid circles* represent observed concentrations. *Solid lines* represent fitted kinetic models. *C*, EIC integration of luteolin glycosylation products produced by R462 mutant libraries of UGT95A1 shows that the residue may be replaced by neutral residues to increase the enzyme's preference for 3′-*O* glycosylation. Data are mean ± s.d. of three technical replicates. *D*, kinetic analysis of luteolin glycosylation reaction mediated by UGT95A1 R462A shows the amount of each product observed over time. UGT95A1 R462A produces luteolin-3′-*O*-glucoside at a rate similar to that observed of the WT enzyme but suppresses the production of all other products, resulting in an apparent increase in regioselectivity. Reactions consisting of 250 nM enzyme, 50 μM luteolin, and 1.6 mM UDP-glucose were incubated at 25 °C, quenched with ice-cold trichloroacetic acid, and analyzed by LC-MS. *E*, docking UDP-glucose (*blue*) and luteolin (*yellow*) to UGT95A1 using AutoDock Vina 1.2.0 shows that luteolin bind with its 3′-hydroxyl group pointing toward the catalytic histidine residue and its 4′-hydroxyl group coordinated by E347 and R462 residues shown in this study to be important for enzyme regioselectivity. EIC, extracted ion chromatogram.

slower while forming the major product at a rate 20 to 85% faster than the WT enzyme (Figs. 4*D*, S6, and Table S2). These results suggest that R462 residue plays a role in the formation of 7-*O* and 4′-*O* glycosylation products.

To determine the function of E347 and R462 in regioselectivity, we docked luteolin and UDP-glucose onto the structural model of UGT95A1. The docking model (File S2) shows both substrates binding in positions comparable to

those in doubly bound crystal structures of related family 1 GT enzymes (PDB ID: 2C1Z, 2VCE, 6JEM) (Figs. 4*E* and S7). Luteolin and UDP-glucose interact predominantly with the N-terminal and the C-terminal domains, respectively. Luteolin interacts with the enzyme using its multiple hydroxyl groups and polar interactions. Notably, luteolin is found with its 3′-*O* pointing toward the catalytic histidine residue (H38) and its 4′-*O* coordinated by E347 and R462 residues, consistent with UGT95A1 primarily producing 3′-*O* glycosylation product (Figs. 4*E* and S7). UDP-Glu binds to the enzyme by interacting with residues in the PSPG box including the central HCGWNS motif (H440 to S445). Notably, D464 is observed interacting with the glucose moiety of UDP-glucose, supporting our biochemical finding that the residue is important for overall UGT95A1 activity. When docking was performed on E347L mutant, luteolin was found to bind with its 7-*O* pointing toward the catalytic residue instead, and R462 was observed near the 2-hydroxyl group of glucose in UDP-glucose (Fig. S8 and File S2). Together, these observations support our biochemical results that E347 and R462 are crucial for orienting luteolin in the correct binding conformation for selective 3′-*O* glycosylation.

Systematic mutational and kinetics studies performed here revealed that, while the substrate binding cleft is surrounded by many amino acids, only a few of them play key roles in either enzyme overall activity or regiospecificity. Most mutations of the active site residues were well tolerated and yielded little to no detectable change in the selective 3′-*O* luteolin glycosylation activity of UGT95A1, suggesting that their catalytic contributions were minor. We found however that four residues, namely M145, E347, R462, and D464, were indispensable for the observed behavior of UGT95A1. Notably, E347 was determined to be crucial for the enzyme's strong regioselectivity for the 3′-*O* of luteolin, potentially by coordinating the 4′-*O* of luteolin and orienting it optimally for the desired reaction, while R462 residue could be mutated to increase the regiospecificity profile of the enzyme.

Distal DRs may be drastically altered without affecting catalysis

To determine the prevalence of 3′-*O* selectivity, we investigated the activity of three close homologs of UGT95A1. *Aa*UGT95 (GenBank: PWA95780.1; 72.5% identity), *Cc*UGT95 (NCBI Reference Sequence: XP_024995450.1; 75.65% identity) and *Mm*UGT95 (GenBank: KAD2002359.1; 73.38% identity) from *Artemisia annua*, *Cynara cardunculus* var. scolymus, and *Mikania micrantha*, respectively, were cloned, expressed, purified (Fig. S1), and tested for their activity on luteolin and UDP-glucose. All three enzymes exhibited glycosylation activity on luteolin and produce luteolin-3′-*O*-glucoside as the major product, similarly to UGT95A1 (Fig. 5*A*)

To determine the basis of specificity of the homologous enzymes, we examined their sequences and found that the three enzymes possess all residues homologous to those identified above as important for enzymatic activity and regiospecificity of UGT95A1 (Fig. 6). Indeed, alanine mutation of residues homologous to E347 in *Aa*UGT95 (E352) and *Mm*UGT95 (E312) resulted in enzymes favoring 7-*O* glycosylation product (Fig. S9), suggesting that they all share a similar mechanism to attain regioselectivity. Interestingly, the four enzymes differ greatly in the three DRs found in the N-terminal domain (DR1-3), both in length and composition, suggesting that these DRs play little or no significant role in regioselectivity and catalysis.

To test this hypothesis and investigate the role of the DRs, we designed a series of deletion mutants of UGT95A1 lacking one or more DRs (Fig. 5*B*) and investigated their activity on luteolin. Individual and combined deletions of the DRs in the N-terminal domain (DR1-3; Del 1–7 mutants) had little effect on enzymatic activity compared to WT (Fig. 5*C*). In fact, kinetic analysis showed that deleting all of DR1-3 (Del 14) resulted in an enzyme that is slightly faster and more specific for 3′-*O* position than UGT95A1 (Fig. S10 and Table S2). This result, along with the previous observation from molecular dynamics simulations that DR1-3 regions extend far from the active site, suggests that they may have little to no role in catalysis.

On the other hand, many deletion variants lacking parts of the DRs in the C-terminal domain (DR4-6; Del 8–13 mutants) showed compromised activity or regiospecificity. Deletions of DR4 (Del 11–12) abolished activity entirely, while the deletion of DR6 (Del 13) severely affected activity (Fig. 5*C*). Interestingly, DR5 could be significantly shortened by deleting residues 393 to 398 without affecting efficiency (Del 8; Fig. 5*C*); however, deleting residues 380 to 392 did affect overall activity and regioselectivity (Del 9,10; Fig. 5*C*). Kinetic study of Del 9 showed that its rate of 3′-*O* glycosylation was reduced by 20-fold while that of 7-*O* glycosylation was reduced 3 to 5 fold compared to the WT enzyme, bringing the two rates to the same order of magnitude (Figs. S11, S12, and Table S2). Visual investigation of structural models and molecular dynamic trajectories suggests that F383 in DR5 potentially interacts with F32, F33, and L114 to make up part of the hydrophobic core of the N-terminal domain. Although regions analogous to DR4-6 are frequently not observed in the crystal structures of family 1 GT enzymes (38), DR4-6 are mostly conserved among the enzyme homologs tested in this study, consistent with the notion that conserved residues usually correspond to functionally important sites. The precise roles of these regions in family 1 GT function require further investigation.

Discussion

This structural and biochemical study of the glycosyltransferase UGT95A1 from *P. officinarum* dissects apart the uncommon regiospecificity for the 3′-hydroxyl group of a flavone sugar acceptor substrate luteolin. The obtained crystal structure of UGT95A1 revealed that the enzyme adopts the GT-B fold, similarly to previously reported plant family 1 GT structures. The reported structure represents the second crystal structure of family 1 GT enzymes with a preference for the 3′-hydroxyl group of the sugar acceptor substrate. The first reported structure of a family 1 GT exhibiting a preference for 3′-*O*-glycosylation is UGT71G1 from *Medicago truncatula*

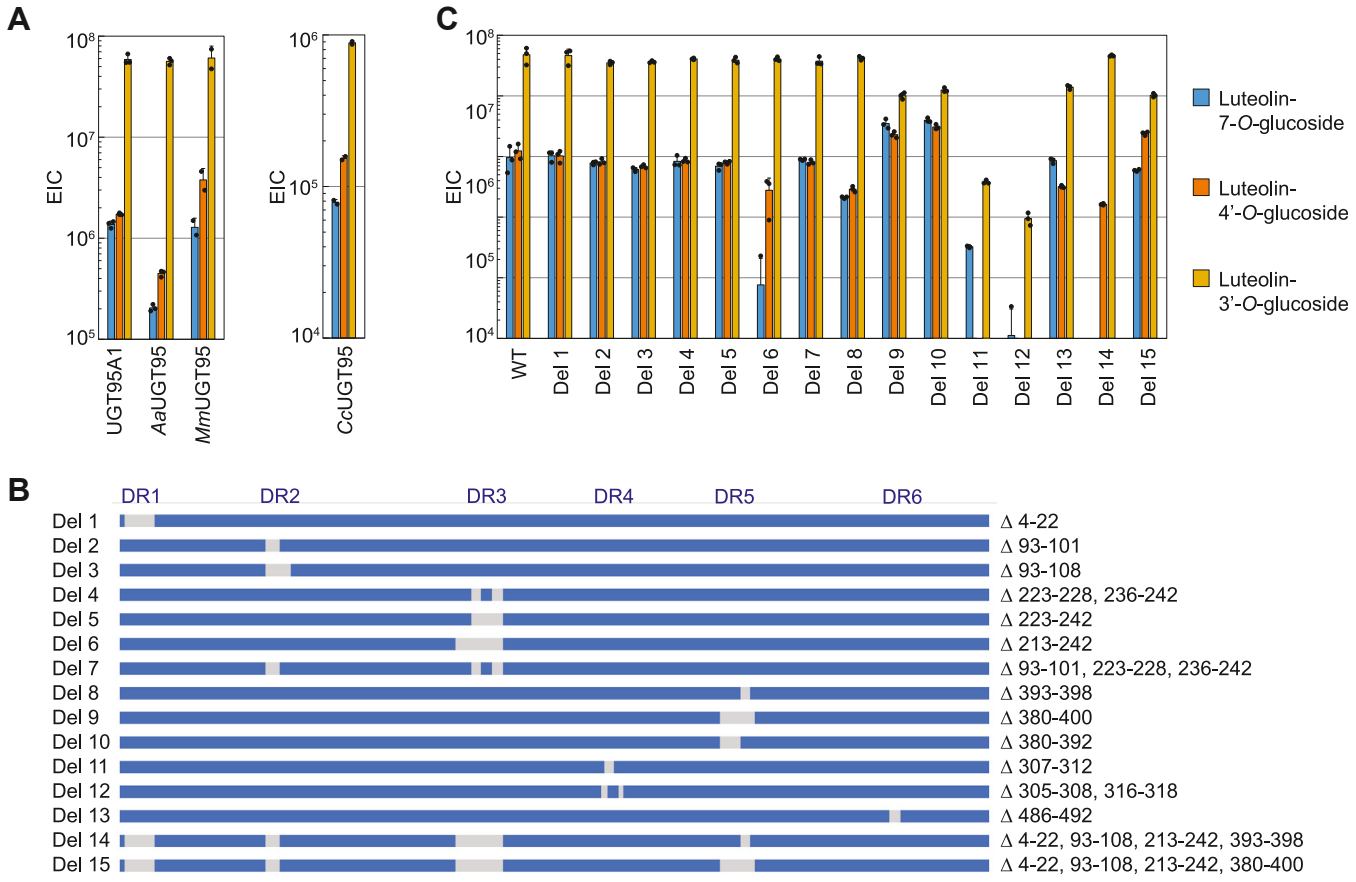

Figure 5. Contributions of distal disordered regions to UGT95A1 enzymatic activity and regioselectivity determined by analysis of homologs and deletion mutants. *A*, EIC integration of luteolin glycosylation products produced by close homologs of UGT95A1, *Aa*UGT95, *Cc*UGT95, and *Mm*UGT95 shows that all three enzymes exhibit a strong preference for 3′-O-glycosylation, similarly to UGT95A1. While *Aa*UGT95 and *Mm*UGT95 produce similar levels of products to UGT95A1, *Cc*UGT95 produces much weaker product signals and are shown on a separate y-axis although data were collected in the same setting. Data are mean ± s.d. of three technical replicates. *B*, diagram depicts disordered regions deleted in deletion mutants tested in this study. Proteins are depicted from N terminus (*left*) to C terminus (*right*) in *blue rectangles*, and deleted residues are represented in *gray*. *C*, EIC integration of luteolin glycosylation products produced by UGT95A1 deletion mutants shows that while deletions of disordered regions (DRs) in the N-terminal domain result in enzymes with WT-like behaviors, deletions of DRs in the C-terminal domain severely affect enzyme activity profile. Data are mean ± s.d. of three technical replicates. EIC, extracted ion chromatogram.

(39). The authors identified residues that are now known as the conserved catalytic dyad of plant family 1 GT enzymes but did not report the basis of substrate regioselectivity.

By investigating the obtained crystal structure and performing mutational biochemical studies, we determined the structural components implicated in UGT95A1's efficiency and regiospecificity. While many point mutations near the substrate binding site are well tolerated, changes to residues M145 and D464 severely impacted enzyme activity. We found E347 to be important for maintaining a strong preference for 3′-*O* glycosylation and R462 to promote 7-*O* and 4′-*O* glycosylation. These findings were further supported by kinetic and docking studies. Our study represents the first time structural factors responsible for the 3′-*O* flavonoid glycosylation regioselectivity have been experimentally determined, and the first time the regioselectivity of glycosyltransferases has been quantified *via* kinetic analysis.

The unique regioselectivity and the underlying structural determinants were further confirmed in homologous enzymes *Aa*UGT95, *Cc*UGT95, and *Mm*UGT95 from *A. annua*, *C. carduncul*us var. scolymus, and *M. micrantha*, respectively. The result suggests that a strong preference for 3′-*O* glycosylation enabled by a properly positioned glutamate residue (E347 in UGT95A1; E352 in *Aa*UGT95; E312 in *Mm*UGT95) may be widespread among UGT95 or clade Q enzymes. This result may be used to further infer regioselectivity of other related enzymes in the subfamily. However, a methionine residue is found in the equivalent position in UGT71G1, suggesting that UGT71G1 uses a different mechanism to exert regioselectivity.

Beside the role of binding site residues, this report also studied DRs for their roles in catalysis. The crystal structure of UGT95A1 consists of six unusually large (up to 36 amino acids) unresolved regions—three in the N-terminal domain and three in the C-terminal domain. Structural modeling, molecular dynamics, and SAXS were used to investigate these regions and showed that they were highly dynamic in relation to the structured protein core, suggesting that they are disordered loops. The simulations further showed the loops are in partially extended conformations, interacting with one another and with the structured part of the protein *via* side-chain charge-charge interactions. By comparing homologous

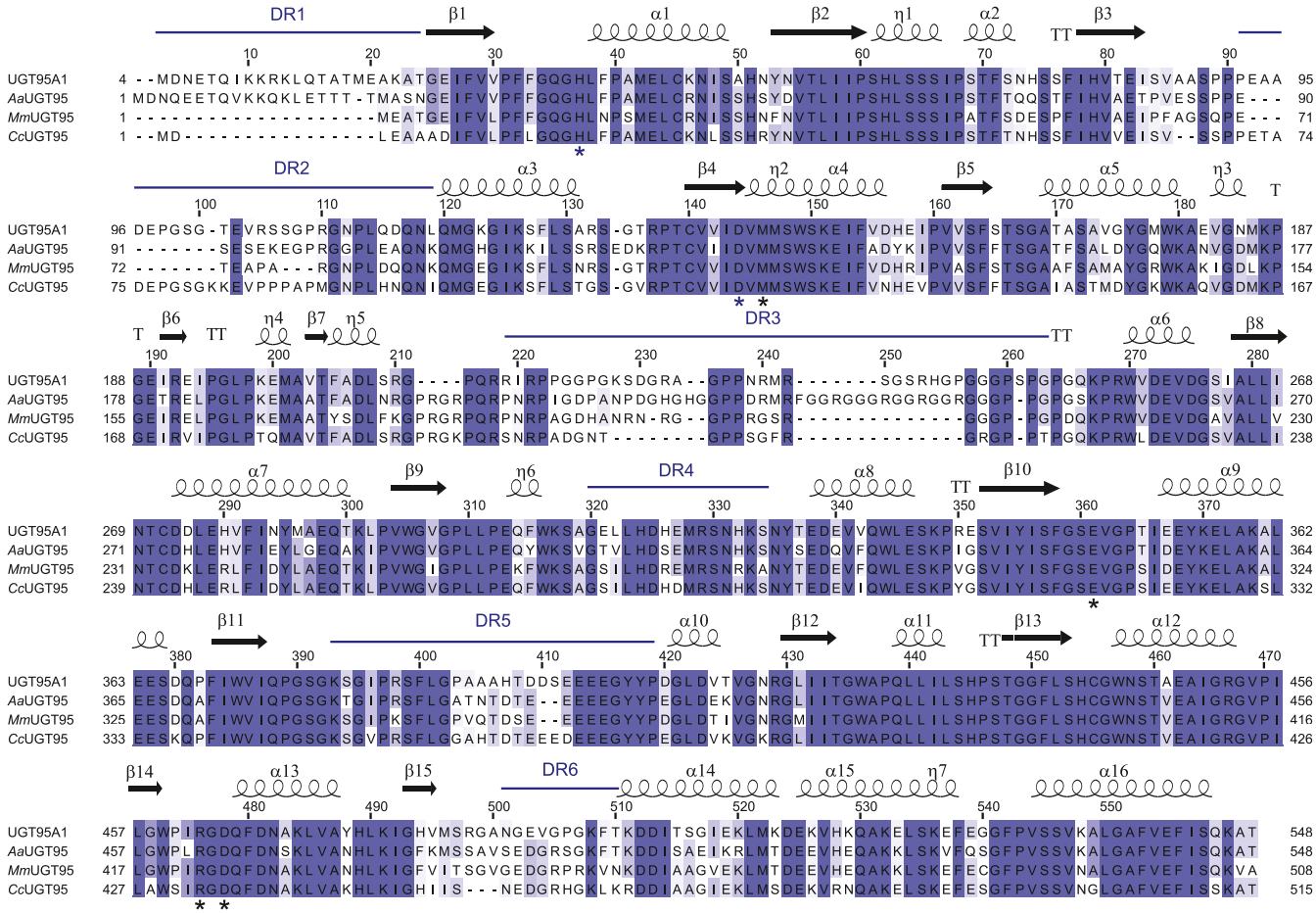

Figure 6. Multiple sequence alignment of UGT95A1 and its homologs *Aa*UGT95, *Cc*UGT95, and *Mm*UGT95. Secondary structural features and disordered regions (DRs) derived from UGT95A1 crystal structure are depicted above the corresponding amino acid residues. Catalytic residues are marked with *blue asterisks* below the amino acid residues. Key residues for activity and regioselectivity determined in this study are marked with *black asterisks*.

enzymes and deletion mutations, we showed that DRs in the N-terminal domain were not required for catalysis *in vitro*. On the other hand, DRs in the C-terminal domain, which are mostly conserved among all the homologs in this study, were required for proper enzymatic function. The DRs that appear dispensable in this study may instead have a function *in vivo* not directly measured here. They may regulate the protein through allostery, posttranslational modifications, or interactions with other cellular factors, as has been shown for DRs found in other enzymes. They may also contribute to protein stability. The precise role of these regions requires further investigations.

While the relationship between structural features and enzymatic activity are necessarily complex, our study showed that a thorough investigation of individual parts may be used to understand the underlying mechanism of efficiency and regioselectivity of family 1 GTs.

Experimental procedures

Bacterial strain

E. coli XL1 Blue was used for DNA construction. *E. coli* BL21*(DE3)-T1^{R} was used for heterologous protein expression.

Gene and plasmid construction

Standard molecular biology techniques were used to carry out plasmid construction. All PCR amplifications were carried out with Phusion High-Fidelity DNA polymerase with primer annealing temperatures 4 to 8 °C below the T_m. DNA assembly was performed using NEBuilder according to the manufacturer's recommended protocol. For analysis and isolation of DNA fragments from agarose gel, 1% gel and QIAquick PCR Purification Kit (Qiagen) were used. All constructs were verified by sequencing (UC Berkeley DNA Sequencing Facility and Genewiz from Azenta Life Sciences). Primers used in this study are shown in Table S1.

For the construction of GT enzyme expression plasmids pGEX-6P-1-GT, coding sequences for GT enzymes were amplified from synthetic gene blocks using primers [enzyme name]_F and [enzyme name]_R. Gene blocks used in this study are shown in Table S1. GST-UGT fusion constructs were cloned by assembling PCR products into the BamHI site of pGEX-6P-1 *via* Gibson assembly. *E. coli* cells transformed with the assembly mixture were selected on culture medium containing carbenicillin.

Each point mutation of UGT95A1 was introduced with two sets of primers that amplify two overlapping DNA pieces. One fragment encoding protein sequence upstream of the mutation was amplified with the forward primer 95A1Mutants_F1 and various reverse primers [mutation]_R1 according to the desired mutation. The second fragment encoding protein sequence downstream of the mutation was amplified with various forward primers [mutation]_F2 according to the desired mutation and the reverse primer 95A1Mutnats_R2. PCR products were assembled into the BamHI site of pGEX-6P-1 *via* Gibson assembly.

Deletion mutants were introduced with overlapping DNA pieces amplified with primers bridging new junctions. The DNA piece upstream of the deletion site was amplified with the forward primer 95A1Mutants_F1 and various reverse primers [deletion]_R1 according to the desired deletion. The second DNA piece downstream of the deletion site was amplified with [deletion]_F2 forward primers and 95A1Mutnats_R2 reverse primer. For deletion mutants with multiple deletion sites, primers used to construct single deletion site mutants were combined to amplify appropriate DNA pieces. PCR products were assembled into the BamHI site of pGEX-6P-1 *via* Gibson assembly.

Expression of GST-tagged proteins

Plasmids encoding the proteins of interest were transformed into *E.coli* BL21*(DE3)-T1^{R}. Terrific broth culture medium (0.5 L) with appropriate antibiotic (carbenicillin: 100 μg/ml) in a 2.5 L ultra-yield flask was inoculated with overnight culture of freshly transformed *E. coli* cells. Cells were grown at 37 °C with shaking at 200 rpm to A_{600} = 0.8 to 1.2, at which time, they were cold-shocked on ice for 20 to 40 min. Expression was then induced by addition of 0.5 mM of IPTG. Induced cells were grown at 16 °C with shaking at 200 rpm overnight and harvested by centrifugation at 10,000*g* for 5 min at 4 °C. Cell pellets were stored at −80 °C until purification.

Purification of GST-tagged proteins

Cell pellets were resuspended in Lysis Buffer A (500 mM sodium phosphate, 150 mM sodium chloride, pH 7.5) containing lysozyme (0.5–1 mg/ml) at 5 to 8 ml/g of cell pellet. Cell suspensions were then sonicated using 5 s ON/25 s OFF cycle for a total ON time of 2 min. Cleared cell lysates were then obtained after centrifugation at 14,000*g* for 20 min at 4 °C. Cleared cell lysates were incubated with 1 to 2 ml of Glutathione Sepharose 4B (Cytiva) for 1 h at 4 °C before loading onto the column by gravity flow. The column was washed with 5 to 15 ml of Lysis Buffer A and with Wash Buffer B (50 mM Tris, 150 mM sodium chloride, pH 8.0) until the eluate was negative for protein content when tested by Bradford protein assay reagent (Bio-Rad). PreScission protease was added to the column and incubated overnight. On the following day, protein was eluted from the column with Wash Buffer B. Eluted protein was concentrated using a 30 kDa molecular weight cutoff (MWCO) Amicon Ultra spin concentrator (Millipore).

For *Cc*UGT95 which was purified in GST-fused form, protein was eluted from column with 20 ml of Elution Buffer C (50 mM Tris, 10 mM glutathione, pH 8.0) immediately after washing with Wash Buffer B. Protein was then concentrated using a 30 kDa MWCO Amicon Ultra spin concentrator (Millipore), before being exchanged into Wash Buffer B (50 mM Tris, 150 mM sodium chloride, pH 8.0) using a PD-10 desalting column containing Sephadex G-25 resin (GE HealthCare).

Protein concentration was determined using A280 and protein extinction coefficients calculated from the sequence using the ExPASy ProtParam program. For enzyme assays and SEC-MALS-SAX experiments, purified protein was aliquoted and flash-frozen in liquid nitrogen before storing at −80 °C. Column was cleaned with 30 ml of Elution Buffer C (50 mM Tris, 10 mM glutathione, pH 8.0) and reequilibrated in 30 ml of Lysis Buffer A before the next use. All proteins were analyzed for purity by SDS-PAGE using Acqua Stain Coomassie gel stain.

For crystallography, concentrated protein was filtered through 0.2 μm surfactant-free cellulose acetate filter before loaded onto HiPrep 26/60 Sephacryl S-200 HR connected to the AKTA explorer. Protein was eluted with Size-Exclusion Buffer D (20 mM Tris, 50 mM NaCl, pH 8.0) at 1 ml/min. Fractions with A280 signal were pooled and concentrated using a 30 kDa MWCO Amicon Ultra spin concentrator to 10 mg/ml.

Crystallization, X-ray data collection, and structure determination

UGT95A1 was screened against the following crystallization screens: MCSG-1 (Anatrace); PEG/Ion, Index, and PEGRx (Hampton Research); ShotGun (Molecular Dimensions); and Berkeley Screen (40). Crystals of UGT95A1 were found in Berkeley D6 (0.2 M ammonium acetate, 0.1 M Mes pH 5.5 and 25% PEG 3350). Crystals were then placed in a reservoir solution containing 20% (v/v) glycerol and flash-cooled in liquid nitrogen.

The X-ray dataset for UGT95A1 was collected at the Berkeley Center for Structural Biology beamline 8.2.2 at the Advanced Light Source at Lawrence Berkeley National Laboratory. The diffraction data were processed using the program Xia2 (41). The crystal structure was solved using molecular replacement with the program Phaser (https://www.phaser.cimr.cam.ac.uk/index.php/Phaser_Crystallographic_Software) (42). The atomic positions obtained from the molecular replacement were used to initiate model building using phenix.autobuild within the Phenix suite (43, 44). Structure refinement was performed using the phenix.refine program (45). Manual rebuilding was done using COOT (https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/) (46). RMSD differences from ideal geometries for bond lengths, angles, and dihedrals were calculated with Phenix (44). The stereochemical quality of the final model was assessed by the program MolProbity (47). A summary of crystal parameters, data collection, and refinement statistics can be found in Table S3. Structure and coordinates for UGT95A1 can be found in the

PDB under accession 9BCM. Topological diagram of UGT95A1 based on crystal structure was constructed using PDBSum Generate (48).

Structure modeling

The complete structure of UGT95A1 was modeled using ColabFold v1.5.5: AlphaFold2 using MMseqs2 (https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/batch/AlphaFold2_batch.ipynb) (49) and default parameters (num_relax: 5; template_mode: none; msa_mode: mmseqs2_uniref_env; pair_mode: unpaired_paired; model_type: auto; num_recycles: 3; recycle_early_stop_tolerance: auto; relax_max_iterations: 200; pairing_strategy: greedy; max_msa: auto; num_seeds: 1). The resulting models were visualized by PyMOL 2.5.5 (https://www.pymol.org/).

Molecular dynamics

Full-atom, explicit solvent, molecular dynamics simulations were conducted starting from selected structures of UGT95A1 generated by AlphaFold2. The molecular dynamics simulations were conducted using the software Amber22 (https://ambermd.org/index.php) (30). The explicit solvent was prepared using the module "tleap" in AmberTools22 to be an octahedral solvent box with a solvent buffer of 10 Å around the protein and compensating charges with Na^+ ions. The protein was parameterized using the ff19SB Amber forcefield (31), and the optical proximity correction model was used to describe the water molecules. The energy of the system was minimized for 2000 steps using a sequence of steepest descent and a conjugate gradient algorithm. The system was then heated up from 0 to 300 K in 10,000 steps (2 fs per step) using a Langevin thermostat. A production simulation was then run for 1000 ns (2 fs per step) at 300 K and with the Berendsen barostat for constant pressure. Production runs were computed on the GPU nodes at the UCSC Hummingbird Computational Cluster. CCPTRAJ software (https://amberhub.chpc.utah.edu/cpptraj/) (50) was used to convert from the NetCDF trajectory format to other formats. Results were analyzed using VMD 1.9.3 (https://www.ks.uiuc.edu/Research/vmd/vmd-1.9.3/) (51) and MDTraj 1.9.9 (https://pypi.org/project/mdtraj/) (52). MDTraj KMeans algorithm was applied to cluster each 1000 ns trajectory into 30 clusters and extract a representative (centroid) from each cluster.

Docking

Multiligand molecular docking was conducted on selected protein structural models using the software AutoDock Vina 1.2.0 (https://vina.scripps.edu/) (53, 54). Pdbqt files were prepared using the software OpenBabel 2.4.1 (https://bioweb.pasteur.fr/packages/pack@openbabel@2.4.1) and considering a pH of 7.4. The search space was defined by manually specifying a 3D point within the enzyme active site as the center of a 30 Å cubic box. The default grid spacing of 0.375 Å was used. A very high exhaustiveness setting of 128 was used to increase the number of searches in the optimization. The top scoring poses were saved and manually inspected using PyMOL 2.5.5. The 2D ligand-protein interaction diagrams were constructed using LigPlot + v.2.2 (55).

Size exclusion chromatography coupled small-angle x-ray scattering with in-line multiangle light scattering experiments

SEC-SAXS-MALS data were collected at the Advanced Light Source beamline 12.3.1 at Lawrence Berkeley National Lab (32). The X-ray wavelength was set at λ = 1.24 Å, and the sample-to-detector distance was 2075 mm resulting in scattering vectors (q) ranging from 0.01 to 0.46 $Å^{-1}$. The scattering vector is defined as $q = 4\pi\sin\theta/\lambda$, where 2θ is the scattering angle. Data were collected using a Pilatus 3 × 2 M Detector (Dectris). Normalization and integration of each image were processed as previously described (56). SEC was performed using the 1290 Infinity High-Performance Liquid Chromatography System (Agilent Technologies) coupled to a Shodex KW-802.5 column (Showa Denko). The column was equilibrated with a running buffer of 100 mM Tris pH 7.6 with or without 100 µM UDP-glucose at a flow rate of 0.65 ml/min. The sample (90 µl) was separated by SEC, and the elution was monitored at 280 and 260 nm by an in-line variable wavelength detector (Agilent Technologies). MALS experiments were performed using an in-line 18-angle DAWN HELEOS II light scattering detector connected in tandem to an Optilab differential refractive index detector (Wyatt Technology). System normalization and calibration were performed with bovine serum albumin using a 50 µl sample at 7 mg/ml in the same running buffer.

The light scattering experiments were used to determine the molecular weight across the principal peaks in the SEC analysis. Ultraviolet, MALS, and differential refractive index data were analyzed using Wyatt Astra 7 software (https://www.wyatt.com/products/software/astra.html) to monitor the homogeneity of the sample across the elution peak complementary to the SEC-SAXS signal validation. A purpose-built SAXS flow cell was connected in-line immediately following the complementary spectroscopic techniques, and 2-s X-ray exposures were collected continuously over the 25-min elution. The SAXS frames recorded before the protein elution peak were used to subtract all other frames. The subtracted frames were investigated by radius of gyration (Rg) derived by the Guinier approximation, $I(q) = I(0)\exp(-q2Rg2/3)$ with the limits $qRg < 1.5$. The elution peak was mapped by comparing integral ratios to background and Rg relative to the recorded frame using the program RAW (57). Uniform Rg values across an elution peak represent a homogenous assembly and were merged to reduce noise in the curve. Final merged SAXS profiles were used for further analysis including the Guinier plot that determined aggregation-free state.

The experimental SAXS profiles were deposited to the Simple Scattering database with access code XSEXBUSH. They were then compared to theoretical scattering curves generated from 30 structural clusters generated from molecular dynamic simulations using FoXS (33, 34).

In vitro *enzyme assay for UDP-dependent glycosylation*

All assay mixtures (50 µl) contained 100 mM Tris, 500 µM UDP-glucose, 50 µM substrate, and 1 µM enzyme at pH 7.6. Reactions were incubated at 25 °C for 1 h. Reactions were quenched with 2 µl of 1M trichloroacetic acid, and 3 µl of 50 uM quercetin was added as internal standard. The reactions were then extracted with 550 µl of ethyl acetate and vortexed for 15 to 30 s. The organic layer (250 µl) was transferred and dried using CentriVap Concentrator (Labconco). Samples were kept at −80 °C until analysis.

Identification and quantification of glycosylation products by LC-MS

Samples were resuspended in 50 µl of 50% methanol in water and filtered through a 96-well 0.2 µM polyvinylidene fluoride membrane plate by centrifugation at 1000*g* for 5 min. Samples were analyzed by Q Exactive (Thermo Fisher Scientific) equipped with an electrospray ionization source in positive ionization mode. Samples were analyzed on an Acquity UPLC BEH C18 column (2.1 × 100 mm, 1.7 µm particle size) connected to Vanquish UHPLC (Thermo Fisher Scientific) using a gradient from 18 to 30% of acetonitrile with 0.1% formic acid in water with 0.1% formic acid in 6.5 min at flow rate of 0.4 ml/min following an initial hold at 18% acetonitrile for 1 min at 25 °C.

The identities of the glycosides were determined based on literature precedence and commercial standards. UGT95A1 reactions result in three peaks within 5 ppm of $[M + H]^+$ = 449.1078 *m/z* at 2.5 min, 3.9 min, and 4.5 min, which were assigned as luteolin 7-*O* glucoside, luteolin-4′-*O*-glucoside, and luteolin-3′-*O*-glucoside, respectively, based on the order of elution previously reported (58). Luteolin 7-*O* glucoside was also confirmed by comparison to commercial standard, and luteolin-4′-*O*-glucoside by comparison to the major luteolin glycosylation product produced by UGT90A7 (25). In addition, the reaction also resulted in one mass peak within 5 ppm of $[M + H]^+$ = 611.1612 *m/z*, which is assigned as luteolin-7,3′-di-*O*-glucoside by comparison to commercial standard.

Kinetic characterization of glycosylation activity

Assays were performed in 96-well PCR plates. Assay components were prepared in two solutions in 100 mM Tris pH 7.6. Solution 1 includes 2 mM UDP-Glu; 25, 50, or 75 µM luteolin; and 1.25% dimethyl sulfoxide. Solution 2 includes enzymes at 4-times the final assay concentration. Reactions were initiated by mixing 80 µl of Solution 1 with 20 µl of solution 2 at 25 °C. Aliquots (10 µl) of the reactions were quenched with 20 µl of ice-cold 60 mM trichloroacetic acid at 1, 5, 10, 15, 20, 30, 40, and 60 min after reaction initiation. Reactions were cleaned up by centrifugation at 2.2k rpm for 20 min, and the supernatant was transferred to a new 96-well microplate (Axygen 500 µl, 96-well, V bottom, SBS footprint). Samples were stored at −20 °C until analysis.

Final enzymes and substrate concentrations are as follows. UGT95A1 WT: 125 nM enzyme + 25 µM luteolin; 250 nM enzyme + 50 µM luteolin; and 375 nM enzyme + 75 µM luteolin. UGT95A1 E347L: 750 nM enzyme + 25 µM luteolin; 1500 nM enzyme + 50 µM luteolin; and 2250 nM enzyme + 75 µM luteolin. UGT95A1 R462A: 125 nM enzyme + 25 µM luteolin; 250 nM enzyme + 50 µM luteolin; and 375 nM enzyme + 75 µM luteolin. UGT95A1 Del9: 750 nM enzyme + 25 µM luteolin; 1500 nM enzyme + 50 µM luteolin; 2250 nM enzyme + 75 µM luteolin. UGT95A1 Del14: 75 nM enzyme + 25 µM luteolin; 150 nM enzyme + 50 µM luteolin; 225 nM enzyme + 75 µM luteolin. UGT95A1 Del15: 750 nM enzyme + 25 µM luteolin; 1500 nM enzyme + 50 µM luteolin; 2250 nM enzyme + 75 µM luteolin.

Samples were analyzed by MSD-XT (Agilent Technologies) in positive ionization mode. Samples were analyzed on a Poroshell 120 EC-C8 column (2.1 × 50 mm, 2.7 µm particle size) connected to 1260 Infinity II HPLC (Agilent Technologies) using a gradient from 12.5 to 40% of acetonitrile with 0.1% formic acid in water with 0.1% formic acid in 5.5 min at flow rate of 0.6 ml/min following an initial hold at 12.5% acetonitrile for 1 min at 25 °C. Reactions resulted in three peaks within 0.5 *m/z* of 449.1 at elution times 3.1, 4.0, and 4.4 min, which were designated luteolin-7-*O*-glucoside, luteolin-4′-*O*-glucoside, and luteolin-3′-*O*-glucoside, respectively. Reactions also resulted in a mass peak within 0.5 *m/z* of 611.1 at 2.4, which was designated as luteolin-7,3′-di-*O*-glucoside. Products were monitored at 356 nm, which is the absorption maximum of luteolin (59). Peaks were collected using ChemStation and integrated using OpenLab CDS (https://www.agilent.com/en/product/software-informatics/analytical-software-suite/chromatography-data-systems/openlab-cds) with manual curation. Integration results were used for kinetic modeling.

The reaction measurements were modeled over time by coupling the optimization and the solution of a system of ordinary differential equations describing the mass balances for each species in the system:

$d[\text{luteolin-7-O-glucoside}]/dt = r_1 - r_4$

$d[\text{luteolin-4′-O-glucoside}]/dt = r_2$

$d[\text{luteolin-3′-O-glucoside}]/dt = r_3$

$d[\text{luteolin-7,3′-di-O-glucoside}]/dt = r_4$

where $r_1 = k_1\,[E][S]$; $r_2 = k_2\,[E][S]$; $r_3 = k_3\,[E][S]$; $r_4 = k_4\,[E][\text{luteolin-7-O-glucoside}]$

[E] = enzyme concentration;

[S] = substrate concentration.

= $[S]_{t=0}$ - [luteolin-7-*O*-glucoside] - [luteolin-4′-*O*-glucoside] - [luteolin-3′-*O*-glucoside] - [luteolin-7,3′-di-*O*-glucoside]

and $[\text{luteolin-7-O-glucoside}]_{t=0} = [\text{luteolin-4′-O-glucoside}]_{t=0} = [\text{luteolin-3′-O-glucoside}]_{t=0} = [\text{luteolin-7,3′-di-}O\text{-glucoside}]_{t=0} = 0$

and k_i = the pseudo first order kinetics constant for the formation of the product i.

The model assumes an isothermal, perfectly mixed system and considers the UDP-glucose cosubstrate to be in large excess. The model assumes pseudo first-order kinetics with respect to the luteolin substrate, and that the diglucoside forms by glycosylation of luteolin-7-*O*-glucoside only. The concentrations of all species were calculated from ion counts and the mass balance described above using response factors defined as:

$[c_i]$ = response factor(c_i) × ion counts(c_i)

in which the response factors were calculated by minimizing the objective function:

$$f = \Sigma_i \left(\frac{C_i - ci^{obs}}{\max\left(ci^{obs}\right)} \right)^2$$

To facilitate fitting the parameters, we specified manual ranges for the kinetic constants and conducted a constrained multistart (250 iterations) optimization using MATLAB R2023a ("lsqnonlin" algorithm). As a solver for the system of equations, an adaptive Runge–Kutta algorithm was used ('ode45'). Each set of results for an enzyme and initial substrate concentration were fitted independently.

Data availability

All data are contained within the article.

Supporting information—This article contains supporting information.

Acknowledgments—This work was part of the DOE Joint BioEnergy Institute (https://www.jbei.org) supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, supported by the U.S. Department of Energy, Energy Efficiency and Renewable Energy, Bioenergy Technologies Office, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy. Crystallography and SEC-SAXS-MALS work was conducted at the Advanced Light Source (ALS), a national user facility operated by Lawrence Berkeley National Laboratory on behalf of the Department of Energy, Office of Basic Energy Sciences, through the Integrated Diffraction Analysis Technologies (IDAT) program, supported by DOE Office of Biological and Environmental Research. Additional support comes from the National Institute of Health project ALS-ENABLE (P30 GM124169).

Author contributions—S. S., V. B., Y. F. S., J. H. P., M. H. and C. R. B. investigation; S. S. and P. M. S. writing–original draft; S. S. and P. M. S. writing–review and editing; S. S. conceptualization; V. B., J. H. P., and M. H. formal analysis; Y. F. S. and C. R. B. methodology; P. D. A. and P. M. S. supervision.

Conflict of interest—The authors declare no conflicts of interest with the contents of this article.

Abbreviations—The abbreviations used are: DR, disordered region; GTs, glycosyltransferases; PDB, Protein Data Bank; PSPG, plant secondary product glycosylation; SEC-SAXS-MALS, size-exclusion chromatography coupled with small-angle X-ray scattering and multiangle light scattering.

References

1. Gachon, C. M. M., Langlois-Meurinne, M., and Saindrenan, P. (2005) Plant secondary metabolism glycosyltransferases: the emerging functional analysis. *Trends Plant Sci.* **10**, 542–549
2. Heiling, S., Llorca, L. C., Li, J., Gase, K., Schmidt, A., Schäfer, M., *et al.* (2021) Specific decorations of 17-hydroxygeranyllinalool diterpene glycosides solve the autotoxicity problem of chemical defense in Nicotiana attenuata. *Plant Cell* **33**, 1748–1770
3. Sugimoto, K., Iijima, Y., Takabayashi, J., and Matsui, K. (2021) Processing of airborne green leaf volatiles for their glycosylation in the exposed plants. *Front. Plant Sci.* **12**, 721572
4. Sugimoto, K., Ono, E., Inaba, T., Tsukahara, T., Matsui, K., Horikawa, M., *et al.* (2023) Identification of a tomato UDP-arabinosyltransferase for airborne volatile reception. *Nat. Commun.* **14**, 677
5. Chen, W., Xiao, Z., Wang, Y., Wang, J., Zhai, R., Lin-Wang, K., *et al.* (2021) Competition between anthocyanin and kaempferol glycosides biosynthesis affects pollen tube growth and seed set of Malus. *Hortic. Res.* **8**, 173
6. Gleadow, R. M., and Møller, B. L. (2014) Cyanogenic glycosides: synthesis, physiology, and phenotypic plasticity. *Annu. Rev. Plant Biol.* **65**, 155–185
7. Hansen, C. C., Sørensen, M., Bellucci, M., Brandt, W., Olsen, C. E., Goodger, J. Q. D., *et al.* (2023) Recruitment of distinct UDP-glycosyltransferase families demonstrates dynamic evolution of chemical defense within Eucalyptus L'Hér. *New Phytol.* **237**, 999–1013
8. Cowan, M. F., Blomstedt, C. K., Møller, B. L., Henry, R. J., and Gleadow, R. M. (2021) Variation in production of cyanogenic glucosides during early plant development: a comparison of wild and domesticated sorghum. *Phytochemistry* **184**, 112645
9. Kytidou, K., Artola, M., Overkleeft, H. S., and Aerts, J. M. F. G. (2020) Plant glycosides and glycosidases: a treasure-trove for therapeutics. *Front. Plant Sci.* **11**, 357
10. Patel, S. (2016) Plant-derived cardiac glycosides: role in heart ailments and cancer management. *Biomed. Pharmacother.* **84**, 1036–1041
11. Ceunen, S., and Geuns, J. M. C. (2013) Steviol glycosides: chemical diversity, metabolism, and function. *J. Nat. Prod.* **76**, 1201–1228
12. Brahmachari, G., Mandal, L. C., Roy, R., Mondal, S., and Brahmachari, A. K. (2011) Stevioside and related compounds - molecules of pharmaceutical promise: a critical overview. *Arch. Pharm. (Weinheim)* **344**, 5–19
13. Elekofehinti, O. O., Iwaloye, O., Olawale, F., and Ariyo, E. O. (2021) Saponins in cancer treatment: current progress and future prospects. *Pathophysiology* **28**, 250–272
14. Güçlü-Ustündağ, O., and Mazza, G. (2007) Saponins: properties, applications and processing. *Crit. Rev. Food Sci. Nutr.* **47**, 231–258
15. Sirirungruang, S., Barnum, C. R., Tang, S. N., and Shih, P. M. (2023) Plant glycosyltransferases for expanding bioactive glycoside diversity. *Nat. Prod. Rep.* **40**, 1170–1180
16. Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res.* **37**, D233–D238
17. Louveau, T., and Osbourn, A. (2019) The sweet side of plant-specialized metabolism. *Cold Spring Harb. Perspect. Biol.* https://doi.org/10.1101/cshperspect.a034744
18. Wang, Z.-L., Gao, H.-M., Wang, S., Zhang, M., Chen, K., Zhang, Y.-Q., *et al.* (2020) Dissection of the general two-step di-C-glycosylation pathway for the biosynthesis of (iso)schaftosides in higher plants. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 30816–30823
19. Brazier-Hicks, M., Offen, W. A., Gershater, M. C., Revett, T. J., Lim, E.-K., Bowles, D. J., *et al.* (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
20. Teze, D., Coines, J., Fredslund, F., Dubey, K. D., Bidart, G. N., Adams, P. D., *et al.* (2021) O -/N -/S -specificity in glycosyltransferase catalysis: from mechanistic understanding to engineering. *ACS Catal.* https://doi.org/10.1021/acscatal.0c04171
21. Wilson, A. E., and Tian, L. (2019) Phylogenomic analysis of UDP-dependent glycosyltransferases provides insights into the evolutionary landscape of glycosylation in plant metabolism. *Plant J.* **100**, 1273–1288
22. Xia, E., Tong, W., Hou, Y., An, Y., Chen, L., Wu, Q., *et al.* (2020) The reference genome of tea plant and resequencing of 81 diverse accessions provide insights into its genome evolution and adaptation. *Mol. Plant* **13**, 1013–1026

23. Sun, L., Zhao, L., Huang, H., Zhang, Y., Wang, J., Lu, X., *et al.* (2022) Genome-wide identification, evolution and function analysis of UGTs superfamily in cotton. *Front. Mol. Biosci.* **9**, 965403
24. Yao, Y., Gu, J., Luo, Y., Wang, Y., Pang, Y., Shen, G., *et al.* (2022) Genome-wide analysis of UGT gene family identified key gene for the biosynthesis of bioactive flavonol glycosides in Epimedium pubescens Maxim. *Synth. Syst. Biotechnol.* **7**, 1095–1107
25. Witte, S., Moco, S., Vervoort, J., Matern, U., and Martens, S. (2009) Recombinant expression and functional characterisation of regiospecific flavonoid glucosyltransferases from Hieracium pilosella L. *Planta* **229**, 1135–1146
26. Wilson, A. E., Wu, S., and Tian, L. (2019) PgUGT95B2 preferentially metabolizes flavones/flavonols and has evolved independently from flavone/flavonol UGTs identified in Arabidopsis thaliana. *Phytochemistry* **157**, 184–193
27. Su, X., Shen, G., Di, S., Dixon, R. A., and Pang, Y. (2017) Characterization of UGT716A1 as a multi-substrate UDP:flavonoid glucosyltransferase gene in ginkgo biloba. *Front. Plant Sci.* **8**, 2085
28. Del Giudice, R., Putkaradze, N., Dos Santos, B. M., Hansen, C. C., Crocoll, C., Motawia, M. S., *et al.* (2022) Structure-guided engineering of key amino acids in UGT85B1 controlling substrate and stereo-specificity in aromatic cyanogenic glucoside biosynthesis. *Plant J.* **111**, 1539–1549
29. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589
30. Case, D. A., Aktulga, H. M., Belfon, K., Cerutti, D. S., Cisneros, G. A., Cruzeiro, V. W. D., *et al.* (2021) AmberTools. *J. Chem. Inf. Model.* **63**, 6183–6191
31. Tian, C., Kasavajhala, K., Belfon, K. A. A., Raguette, L., Huang, H., Migues, A. N., *et al.* (2020) ff19SB: amino-acid-specific protein backbone parameters trained against quantum mechanics energy surfaces in solution. *J. Chem. Theor. Comput.* **16**, 528–552
32. Rosenberg, D. J., Hura, G. L., and Hammel, M. (2022) Size exclusion chromatography coupled small angle X-ray scattering with tandem multiangle light scattering at the SIBYLS beamline. *Meth. Enzymol.* **677**, 191–219
33. Schneidman-Duhovny, D., Hammel, M., and Sali, A. (2010) FoXS: a web server for rapid computation and fitting of SAXS profiles. *Nucleic Acids Res.* **38**, W540–W544
34. Schneidman-Duhovny, D., Hammel, M., Tainer, J. A., and Sali, A. (2013) Accurate SAXS profile computation and its assessment by contrast variation experiments. *Biophys. J.* **105**, 962–974
35. Maharjan, R., Fukuda, Y., Nakayama, T., Nakayama, T., Hamada, H., Ozaki, S. I., *et al.* (2020) Crown-ether-mediated crystal structures of the glycosyltransferase PaGT3 from Phytolacca americana. *Acta Crystallogr. D Struct. Biol.* **76**, 521–530
36. Yang, T., Zhang, J., Ke, D., Yang, W., Tang, M., Jiang, J., *et al.* (2019) Hydrophobic recognition allows the glycosyltransferase UGT76G1 to catalyze its substrate in two orientations. *Nat. Commun.* **10**, 3214
37. Offen, W., Martinez-Fleites, C., Yang, M., Kiat-Lim, E., Davis, B. G., Tarling, C. A., *et al.* (2006) Structure of a flavonoid glucosyltransferase reveals the basis for plant natural product modification. *EMBO J.* **25**, 1396–1405
38. Li, L., Modolo, L. V., Escamilla-Trevino, L. L., Achnine, L., Dixon, R. A., and Wang, X. (2007) Crystal structure of Medicago truncatula UGT85H2–insights into the structural basis of a multifunctional (iso) flavonoid glycosyltransferase. *J. Mol. Biol.* **370**, 951–963
39. Shao, H., He, X., Achnine, L., Blount, J. W., Dixon, R. A., and Wang, X. (2005) Crystal structures of a multifunctional triterpene/flavonoid glycosyltransferase from Medicago truncatula. *Plant Cell* **17**, 3141–3154
40. Pereira, J. H., McAndrew, R. P., Tomaleri, G. P., and Adams, P. D. (2017) Berkeley Screen: a set of 96 solutions for general macromolecular crystallization. *J. Appl. Crystallogr.* **50**, 1352–1358
41. Winter, G., Lobley, C. M. C., and Prince, S. M. (2013) Decision making in xia2. *Acta Crystallogr. D Biol. Crystallogr.* **69**, 1260–1273
42. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674
43. Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Zwart, P. H., Hung, L. W., *et al.* (2008) Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr. D Biol. Crystallogr.* **64**, 61–69
44. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221
45. Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., *et al.* (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 352–367
46. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132
47. Williams, C. J., Headd, J. J., Moriarty, N. W., Prisant, M. G., Videau, L. L., Deis, L. N., *et al.* (2018) MolProbity: more and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315
48. Laskowski, R. A., Jabłońska, J., Pravda, L., Vařeková, R. S., and Thornton, J. M. (2018) PDBsum: structural summaries of PDB entries. *Protein Sci.* **27**, 129–134
49. Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022) ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682
50. Roe, D. R., and Cheatham, T. E. (2013) PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theor. Comput.* **9**, 3084–3095
51. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38
52. McGibbon, R. T., Beauchamp, K. A., Harrigan, M. P., Klein, C., Swails, J. M., Hernández, C. X., *et al.* (2015) Mdtraj: a modern open library for the analysis of molecular dynamics trajectories. *Biophys. J.* **109**, 1528–1532
53. Trott, O., and Olson, A. J. (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
54. Eberhardt, J., Santos-Martins, D., Tillack, A. F., and Forli, S. (2021) Autodock vina 1.2.0: new docking methods, expanded force field, and python bindings. *J. Chem. Inf. Model.* **61**, 3891–3898
55. Laskowski, R. A., and Swindells, M. B. (2011) LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. *J. Chem. Inf. Model.* **51**, 2778–2786
56. Hura, G. L., Menon, A. L., Hammel, M., Rambo, R. P., Poole, F. L., Tsutakawa, S. E., *et al.* (2009) Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). *Nat. Methods* **6**, 606–612
57. Hopkins, J. B. (2024) BioXTAS RAW 2: new developments for a free open-source program for small-angle scattering data reduction and analysis. *J. Appl. Crystallogr.* **57**, 194–208
58. Xu, T., Wang, C., Jiang, S., Yang, T., and Wu, X. (2022) Glycosylation of luteolin in hydrophilic organic solvents and structure-antioxidant relationships of luteolin glycosides. *RSC Adv.* **12**, 18232–18237
59. Dong, H., Yang, X., He, J., Cai, S., Xiao, K., and Zhu, L. (2017) Enhanced antioxidant activity, antibacterial activity and hypoglycemic effect of luteolin by complexation with manganese(ii) and its inhibition kinetics on xanthine oxidase. *RSC Adv.* **7**, 53385–53395