JBC RESEARCH ARTICLE

Structural and functional studies reveal the molecular basis of substrate promiscuity of a glycosyltransferase originating from a major agricultural pest

Received for publication, September 18, 2023, and in revised form, October 16, 2023 Published, Papers in Press, November 1, 2023,
https://doi.org/10.1016/j.jbc.2023.105421

Ricardo Hernandez Arriaza[1,2], Brendan Abiskaroon[1], Megha Patel[2], Leily Daneshian[2], Anna Kluza[3], Simon Snoeck[4], Maxwell B. Watkins[5], Jesse B. Hopkins[5], Thomas Van Leeuwen[4], Miodrag Grbic[6,7], Vojislava Grbic[6], Tomasz Borowski[3], and Maksymilian Chruszcz[1,2,*]

From the [1]Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, Michigan, USA; [2]Department of Chemistry and Biochemistry, University of South Carolina, Columbia, South Carolina, USA; [3]Jerzy Haber Institute of Catalysis and Surface Chemistry, Polish Academy of Science, Krakow, Poland; [4]Department of Plants and Crops, Faculty of Bioscience Engineering, Ghent University, Ghent, Belgium; [5]The Biophysics Collaborative Access Team (BioCAT), Department of Physics, Illinois Institute of Technology, Chicago, Illinois, USA; [6]Department of Biology, Western University, London, Ontario, Canada; [7]University of La Rioja, Logrono, Spain

Reviewed by members of the JBC Editorial Board. Edited by Robert Haltiwanger

The two-spotted spider mite, *Tetranychus urticae*, is a major cosmopolitan pest that feeds on more than 1100 plant species. Its genome contains an unprecedentedly large number of genes involved in detoxifying and transporting xenobiotics, including 80 genes that code for UDP glycosyltransferases (UGTs). These enzymes were acquired *via* horizontal gene transfer from bacteria after loss in the Chelicerata lineage. UGTs are well-known for their role in phase II metabolism; however, their contribution to host adaptation and acaricide resistance in arthropods, such as *T. urticae*, is not yet resolved. TuUGT202A2 (Tetur22g00270) has been linked to the ability of this pest to adapt to tomato plants. Moreover, it was shown that this enzyme can glycosylate a wide range of flavonoids. To understand this relationship at the molecular level, structural, functional, and computational studies were performed. Structural studies provided specific snapshots of the enzyme in different catalytically relevant stages. The crystal structure of TuUGT202A2 in complex with UDP-glucose was obtained and site-directed mutagenesis paired with molecular dynamic simulations revealed a novel lid-like mechanism involved in the binding of the activated sugar donor. Two additional TuUGT202A2 crystal complexes, UDP-(S)-naringenin and UDP-naringin, demonstrated that this enzyme has a highly plastic and open-ended acceptor-binding site. Overall, this work reveals the molecular basis of substrate promiscuity of TuUGT202A2 and provides novel insights into the structural mechanism of UGTs catalysis.

Glycosylation of small molecules is an essential process of life. When a sugar moiety is attached to endogenous or exogenous small molecules, called acceptors, these molecules' physicochemical and biological properties may be altered (1, 2). This process plays a major role in cellular homeostasis as it regulates the subcellular localization or even secretion of these acceptors (3–5). Glycosyltransferases (GTs) are enzymes, ubiquitous to all forms of life, that catalyze the formation of a glycosidic bond between an activated sugar donor and an acceptor (6). The CAZy database (7), which applies a sequence-based classification, is often used to sort these large number of enzymes. Following this categorization, the GT-1 family of enzymes is of particular interest due to their role in xenobiotic metabolism and defense mechanisms in humans, plants, and insects (4, 8–10). They are well known for their essential role in drug metabolism in the human body and are considered the primary phase II drug-metabolizing enzymes (5, 11–13). GT-1 enzymes are often called UDP glycosyltransferases (UGTs) as they can use several different nucleotide sugars, such as UDP-glucose (UDP-Glc), UDP-galactose, UDP-glucuronic acid, among others (10, 14, 15). They belong to the GT-B superfamily, according to their structural classification (6, 7). The GT-B superfamily is characterized by a highly conserved tertiary structure, despite a low sequence similarity, that consists of two β/α/β Rossmann-like domains facing each other with the catalytic site in the middle (16, 17).

Tetranychus urticae, the two-spotted spider mite (TSSM), is a polyphagous pest that can feed on over 150 crops (18, 19). Such a wide host range indicates this mite's unprecedented xenobiotic responsiveness that is also coincident with their ability to rapidly develop resistance to pesticides (18, 20). Consistently, analysis of genome sequences (21) revealed major expansions of gene families involved in detoxification and transport of xenobiotics. Among them, there are 80 genes encoding for UGTs, proposed to be horizontally transferred from bacteria (22). Snoeck *et al.* (10) were the first ones to show, *in vitro*, the ability of these enzymes to glycosylate a broad range of plant secondary metabolites and acaricides,

* For correspondence: Maksymilian Chruszcz, chruszcz@msu.edu.
Present address for Simon Snoeck: Department of Plant and Microbial Biology, University of Zurich, Zollikerstrasse 107, Zurich CH-8008, Switzerland.

indicating that these UGTs play a role in mite host and pesticide adaptations. A recent eQTL study revealed that the over-expression of at least some UGTs is coregulated with other detoxifying enzymes such as cytochrome P450s, by a master regulator from the HR96 hormone receptor family, reinforcing their role in adaptation and resistance (23).

Flavonoids are plant secondary metabolites that carry a wide range of functions, including signaling, pigmentation, and defense to not only UV radiation but also against herbivores, fungi, and microbes (24–26). In addition, flavonoids have anticancer, antioxidant, and anti-inflammatory properties (27). Furthermore, they are used as preservatives and pigments in food industry (24, 28). It is assumed that specific properties of flavonoids arise from their molecular structure, such as the presence or absence of functional groups, as well as conjugated groups, like sugar moieties (29–31). For example, different flavonoids and their glycoside derivatives possess different anti-herbivore and anti-cancer attributes (23, 32). Consistently, it has been shown that bacteria and herbivores rely on UGTs to detoxify these compounds (10, 33, 34).

In this work, we provide structural and functional characterization of TuUGT202A2 (Tetur22g00270), an enzyme suggested to play a role in the adaptation of *T. urticae* to tomato (10). We report the first crystal structure of a chelicerate UGT, which also represents the first experimental model of the N-terminal end domain of a UGT originated from an animal. The overall fold of the enzyme reveals how evolution has provided TSSM's UGTs with new features, as compared to bacterial UGTs, that are directly linked to its polyphagous ability. Elucidation of the crystal structure of TuUGT202A2 in different catalytically relevant stages allowed to determine some of the conformational changes that drive catalysis. For instance, a novel lid-like mechanism, among the GT-B family, involved in the nucleotide sugar binding is proposed. Co-crystallization of TuUGT202A2 in the presence of UDP and two different substrates, (S)-naringenin and naringin, paired with functional studies, revealed the molecular basis of substrate promiscuity of this enzyme. Due to the folding and catalysis conservation among UGTs across all kingdoms, the results described in this study not only reveal xenobiotics detoxification details, but they provide new insights into the function of UGTs.

Results

Overall structure of TuUGT202A2 apoform and in-solution structural studies via *SEC-SAXS*

The first crystal structure of a UGT originating from a chelicerate (PDB code: 6PNT, $P3_121$ space group) was determined at 1.8 Å resolution. It contained one molecule in the asymmetric unit. Almost the entire length of the protein was modeled in the electron density map obtained (residues 10–437). No electron density was observed for the first nine residues and the purification tag (residues: −36 to 9), indicating a high degree of flexibility. This crystal structure revealed the characteristic folding conservation among enzymes belonging to the GT-B superfamily. These enzymes require two substrates for catalysis; thus, the overall folding of this group of enzymes is usually divided into the N-terminal end domain, which plays a major role in the acceptor binding, and the C-terminal end domain, that is primarily involved in the activated sugar binding (6). In the case of TuUGT202A2, we considered that residues 1–248 form the N-terminal end domain, whereas residues 249–437 are part of the C-terminal end. These two domains are mainly composed by the characteristic β/α/β Rossmann-like fold, which face each other and locate the catalytic site between them (Fig. 1) (6). In addition, these two domains present four main regions that are not part of the Rossmann-like fold (Fig. 1). Two of these regions belong to the N-terminal domain, residues 70–113 and 182–210, and the other two belong to the C-terminal domain, residues 247–262 and 401–437 (Fig. 1).

In addition to crystallographic studies, size-exclusion chromatography-small angle X-ray scattering (SEC-SAXS) was used to obtain a more detailed understanding of the structural behavior of this protein in solution (SASBDB accession code: SASDSH5). This was mainly driven by the fact that crystallization experiments failed to provide any structural information on the first nine residues, and the purification tag of our protein construct and the oligomerization state of UGTs is of particular importance as it can modify the catalytic properties of these enzymes (35, 36). Upon evolving factor analysis (Fig. S1) (37) and fitting of different theoretical models to the experimental SAXS data *via* FoXS and MultiFoXS (38) (Fig. S2), it was found that TuUGT202A2 behaves as a monomer in solution, which was consistent with size-exclusion chromatography and computational analysis of the crystal packing (Table S1).

Comparison of apoform TuUGT202A2 with homologous proteins

UGTs are ubiquitous to all kingdoms. As mentioned previously, these enzymes are known for having a high degree of folding similarity despite their low sequence identity. The crystal structure of apoform TuUGT202A2 was compared against all other structures of PDB *via* DALI and PDBeFold servers, using default parameters for both servers (39, 40). Similar results were obtained within the top ten ranked structures (Tables S2 and S3). The two servers agreed on ranking TuUGT203A3 (PDB code: 7MCO), another UGT originating from *T. urticae*, deposited by our research group, as the most similar structure. The rest of the top ranked structures corresponded to GT-1 glycosyltransferases originated from bacteria: macrolide glycosyltransferases from *Streptomyces antibioticus*, OleI; ginsenoside-biosynthetic UDP-glycosyltransferase from *Bacillus subtilis*, YjiC; and two calicheamicin glycosyltransferases originating from *Escherichia coli*, CalG1 and CalG3. These results are consistent with the idea that spider mite UGTs were horizontally acquired from bacteria (22). None of these homologous proteins have a sequence identity higher than 34% to TuUGT202A2. As expected, a multiple sequence alignment (MSA) between these enzymes shows a higher degree of sequence similarity in the C-

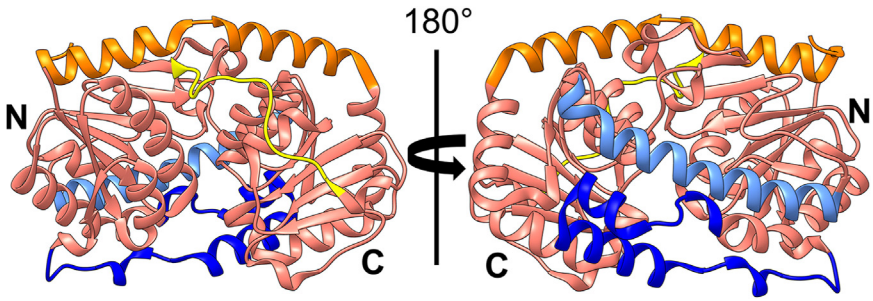

Figure 1. Overall structure of TuUGT202A2. The portions of the N-terminal and C-terminal end domains that are part of the characteristic Rossmann-like fold are shown in *salmon color* and labeled as "N" and "C", respectively. The additional four regions are depicted in four different colors. The two regions that are located within the N-terminal domain are colored *blue* (residues: 70–113) and *light blue* (182–210). The ones belonging to the C-terminal end are depicted in *yellow* (residue: 247–262) and *orange* (residues: 401–437). UGT, UDP glycosyltransferase.

terminal end domain, the sugar-binding domain, than the N-terminal end domain, the acceptor-binding site (Fig. S3).

At the tertiary structure level, when comparing TuUGT202A2 with TuUGT203A3 (PDB code: 7MCO, chain B), the overall folding is very similar (RMSD of 1.0 Å, over 264 Cα atom pairs) (Fig. S4*A*). Even the four additional regions, that are not part of the Rossman-like fold, superimpose well between both structures (Fig. S4*A*). The superimposition of TuUGT202A2 with the structurally similar bacterial glycosyltransferases mentioned above clearly show the high fold conservation of the sugar-binding domain between TSSM and bacteria (Fig. 2). The high degree of similarity expands beyond the Rossmann-like fold portion of this domain and only a few loops are seen to adopt different conformations (Fig. 2). Some of these different conformations, such as the one seen in the loop composed by residues 247–262, in yellow (numbering according to TuUGT202A2), could be attributed to the fact that all these bacterial glycosyltransferase structures were determined in the presence of the activated sugar, which suggests that this loop undergoes conformational changes upon UDP-Glc binding (Fig. 2).

More significant differences in the folding are seen in the acceptor-binding domain (N-terminal domain). This is usually the case when comparing homologous UGTs as it is believed that these differences drive the substrate preferences of each enzyme (17). The TuUGT202A2's Rossmann-like fold portion of the N-terminal end domain superimposes well with the bacterial structures (Fig. 2). However, major differences are seen when comparing the two regions that are not part of this conserved fold (residues 70–113 and 182–210). In TuUGT202A2 structure, these two regions seemed to completely cross to the C-terminal domain; however, nothing similar is seen in the bacterial glycosyltransferases that have their structures deposited in the PDB (Figs. 2 and S4*B*). Due to the relative location of these two regions with respect to the acceptor-binding site, we propose that these differences are product of evolution and distinguish this arthropod UGTs from bacterial glycosyltransferases. This structural analysis suggests that, while UGTs were acquired by TSSM *via* horizontal gene transfer from bacteria, these enzymes have evolved and gained new structural features that most likely play a role in their substrate preferences. These new characteristics could be seen as a "gain of function" for these enzymes that most likely contribute to the broad range of plant secondary metabolites and acaricides TSSM can detoxify.

Currently in the PDB, the only experimental models of UGTs originating from animals are fragments of the C-terminal end domain of three different human UGTs: UGT2B7

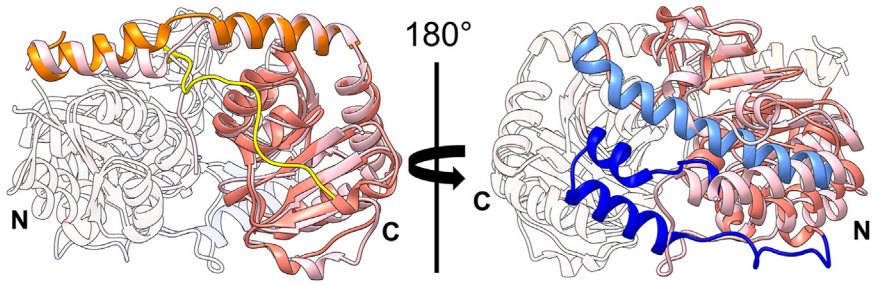

Figure 2. Activated sugar-binding and acceptor-binding domains of TuUGT202A2. Superimposition of apoform TuUGT202A2 (PDB code: 6PNT) with the most similar bacterial glycosyltransferase (OleI, PDB code: 2IYA), according to the Z-score of PDBeFold and DALI servers, exemplify the structural differences in the acceptor-binding domain between UGTs from arthropod and bacteria. TuUGT202A2 is colored following the same scheme as Figure 1, while the ribbon representation of OleI chain B is in *pink*. On the *left*, the ribbon representations of the N-terminal end domains are shown with the highest transparency to allow a clearer visualization of the C-terminal end domain (sugar binding domain). On the *right*, the transparent and colored domains are inverted to ease the visualization of the N-terminal end domain (acceptor-binding domain). The *left picture* clearly demonstrates the folding conservation of the sugar-binding domain across species. All residues superimpose well, except of certain loop regions, such as the loop formed by residues 247–262 (numbering according to TuUGT202A2), *yellow colored*, which can be explained by the fact that this bacterial glycosyltransferase structure was determined in the presence of the activated sugar. On the *right*, the Rossmann-fold–like domain of the acceptor-binding domain superimposes well among these structures, while OleI structures does not present anything like the TuUGT202A2 additional regions that are not part of this conserved fold and cross completely to the N-terminal end domain (colored *blue* and *light blue* in TuUGT202A2). For additional analysis and comparison of TuUGT202A2 with the rest structurally similar bacterial glycosyltransferases, refer to Fig. S4*B*. UGT, UDP glycosyltransferase.

(PDB code: 2O6L), UGT2B10 (PDB code: 7YF5), and UGT2B15 (PDB code: 6IPB). To investigate how similar TuUGT202A2 is to other animal UGTs, we superimposed the crystal structure of the apoform of TuUGT202A2 with the AlphaFold2 models of the full-length human UGTs mentioned before. While it is important to note that human UGTs are known to have a transmembrane domain at the C-terminal end and a signal peptide sequence at the N-terminal end, as compared to TuUGT202A2, the core of these proteins (meaning excluding these two regions) superimpose relatively well (Fig. S5). What is most interesting is that these human UGTs are also predicted to have similar regions that are not part of the Rossmann-like fold in the N-terminal domain but completely cross to the C-terminal domain, as the ones characteristic to TSSM's UGTs (residues 70–113 and 182–210, numbering per TuUGT202A2) (Fig. S5).

Studies of TuUGT202A2 in complex with activated sugar

The crystal structure of TuUGT202A2 in complex with intact UDP-Glc was obtained (Fig. S6). This complex (space group: $P4_12_12$, PDB code: 8SFY), was determined at 2.4 Å and contained only one protein chain in the asymmetric unit. Comparing it to the apoform structure, these two structures are very similar, RMSD of 0.4 Å (between 418 superposed Cα atom pairs). The major difference is seen in the conformation adopted by the loop region of residues 247–262, which is located near UDP-Glc. No electron density was seen for residues 252–256 within this loop, suggesting a high degree of flexibility (Fig. 3*A*). This is consistent with the idea previously presented in section "Comparison of apoform TuUGT202A2 with homologous proteins", which states that this loop undergoes conformational changes upon sugar donor binding (Fig. 3*A*).

From this crystal complex, a detailed description of UDP-glucose–protein interactions can be elucidated (Fig. 3*B*). UDP-Glc is bound in a hydrophilic cavity that was seen open in the apoform crystal structure. It interacts primarily with the C-terminal end of TuUGT202A2 through hydrogen bonds and only one hydrophobic interaction, π stacking between Phe324 and the uracil group of the ligand. Interestingly, this interaction is possible after a significant conformational change on the Phe324 side chain which is seen when comparing UDP-Glc bound and apoform structures. Besides this interaction, the pyrimidine ring interacts with the enzyme through one of the oxo groups. Hydrogen bonds are being formed with the side chain of Ser305 and the backbone carboxyl group of Ile325, which is hydrogen bonding with the pyrimidine nitrogen located between the two oxo groups. Moving upwards on the ligand, the ribose ring forms hydrogen bonds with Glu348, while the α-phosphate group forms four hydrogen bonds, two with Ser345 and two with Asn344, which in both cases include the amino acids' side chain and the amino group from their backbone. On the other hand, the β-phosphate interacts with the side chain of His340 through a bifurcated hydrogen bond. In the case of the glucose moiety, all hydroxyl groups present form hydrogen bonds. The side chains of Asp364 and Gln365 form hydrogen bonds with two consecutive hydroxyl groups of the hexose. Ser343 uses its side chain and amide backbone group to form two hydrogen bonds with the same hydroxyl group from the hexose. The hydroxyl group located at the methyl group of glucose interacts with Gly24, Ser160, and Asn344, forming a total of three hydrogen bonds. MSA between all complete UGT sequences originating from *T. urticae* showed that all residues that directly interact with the ligand are highly conserved (Fig. S7). On the other hand, these residues are only moderately conserved when compared to the structurally similar bacterial glycosyltransferases previously analyzed (Fig. S3). Overall, this mode of binding of the sugar donor, which relies mainly on interactions mediated by hydrogen bonds, resembles other GT-1s originating from different species, such as plants and bacteria (41, 42). In addition to the structural analysis, to better understand how the binding of UDP-Glc to TuUGT202A2 occurs, the thermodynamic properties for this process were elucidated using ITC experiments (Fig. S8). These studies revealed a reaction stoichiometry value (n) of 1.1 ± 0.1 and the dissociation constant (K_d) was found to be 48.9 ± 15.6 μM. This binding is

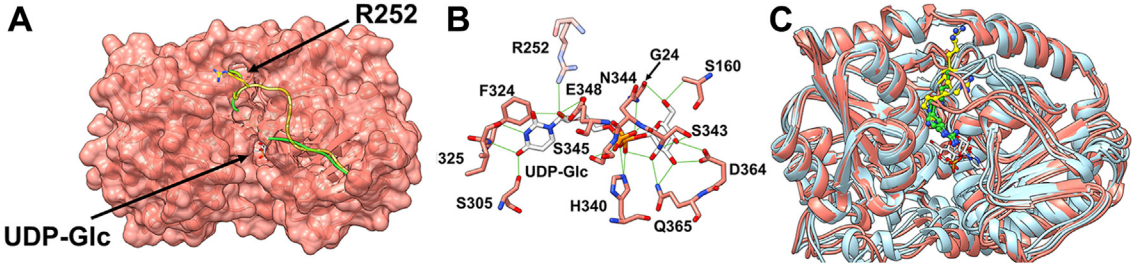

Figure 3. UDP-Glc binding to TuUGT202A2. Structural and computational studies of UDP-Glc (*stick* representation with *light gray* carbon atoms) binding to TuUGT202A2 support the idea of the lid-like mechanism that takes place upon this ligand binding and plays a role in catalysis. *A*, surface representation of TuUGT202A2–UDP-Glc complex (PDB: 8SFY). The loop comprised by residues 247–262 of this structure is shown as *green ribbon* and the residues that could not be modeled in the crystal structure are represented as a *dashed line*. The same loop from the apoform crystal structure (PDB: 6PNT) was superimposed (*yellow ribbon*) to highlight the change in conformation from both structures. Arg252 is in *stick* representation. *B*, atomic representation of the interactions observed between UDP-Glc and TuUGT202A2 in PDB code: 8SFY. This figure includes the putative interactions with Arg252 obtained by superimposing of the UDP-(S)-naringenin crystal structure (PDB code: 8GKN chain A). The observed hydrogen bonds between UDP-Glc and TuUGT202A2 residues are colored *green*. *C*, ribbon representation of the three representative models from the three independent MD simulation of the apoform (*blue colored*) and UDP-Glc (*salmon colored*). Arg252 is in ball-stick representation. *Yellow* is used for the Arg252 in the apoform, while *green* is used to mark Arg252 for complex with UDP-Glc. This loop maintains a "closed" conformation when UDP-Glc is present but moves freely (without closing the hydrophilic-binding site) if the ligand is absent. UDP-Glc, UDP-Glucose; UGT, UDP glycosyltransferase.

mainly enthalpy-driven, shown by a large negative change in enthalpy (ΔH) −24.2 ± 1.9 kJ mol^{-1}. A small positive change in entropy (ΔS) 1.6 ± 1.5 J mol^{-1} K^{-1} indicates entropy's minor, yet favorable, role in this process. Also, the negative Gibbs free energy (ΔG) of −24.6 kJ mol^{-1}, calculated using these parameters, revealed that the association of donor-sugar and TuUGT202A2 takes place spontaneously. These thermodynamic studies are consistent with the binding mode observed in the crystal structures in which noncovalent interactions predominate (43).

In the quest to investigate whether the change in conformation of the loop (residues 247–262) observed in the UDP-Glc complex crystal structure *versus* the apoform is of biological relevance, an Arg252Ala mutant was generated. This mutation was chosen because Arg252 could not be modeled in the crystal structure of UDP-Glc; however, Arg252 forms a hydrogen bond with UDP in the crystal complex that resembles a catalytically productive complex (see section: Structural and functional studies of TuUGT202A2 interactions with naringenin). Kinetic studies performed at saturated concentration of the model substrate, 1-naphthol, and varying the concentration of UDP-Glc revealed that this mutation results in a reduction of the affinity for UDP-Glc and catalytic efficiency, seen by the increase of K_M as compared to the WT enzyme and the decrease on the turnover number (k_{cat}) (Table 1). This demonstrated that Arg252–UDP-Glc interaction plays an important ancillary role in binding and potentially activation of UDP-Glc during catalysis. On the other hand, when UDP-Glc was present in the reaction at saturated levels and the concentration of 1-naphthol was varied, no significant change was seen in the affinity for the acceptor, yet a lower catalytic efficiency was still observed (Table 1). These results implied that Arg252 affects the catalytic process in an additional manner that is independent of the UDP-Glc binding. The role of a flexible loop that acts as a lid, upon the binding of a sugar donor, to protect the catalytic complex, has already been reported for other GTs but not for UGTs (44, 45). To test whether this loop is acting in a protective lid-like manner, we performed molecular dynamic (MD) simulations with the apoform and the UDP-Glc complex. We hypothesized that if there is a lid-like mechanism that requires UDP-Glc to be bound, the flexibility of this loop will decrease in the presence of this nucleotide sugar, whereas in the apoform, the loop will move freely without closing this readily open hydrophilic cavity. Indeed, our computational studies supported this hypothesis. When comparing the representative structures of three independent MD runs for both the apoform and activated sugar complex, this loop stays in an "open" conformation and Arg252 is seen in three different conformations in the apoform. In contrast, when UDP-Glc is present in the hydrophilic cavity, this loop retains this "closed" conformation and Arg252's side chain maintains a conformation that allows it to interact with UDP-Glc in all three representative models (Fig. 3*C*).

Structural and functional studies of TuUGT202A2 interactions with naringenin

The first *in vitro* substrate screening in which TuUGT202A2's activity was tested towards a broad range of plant secondary metabolites and acaricides highlighted the promiscuity of this enzyme towards flavonoids, which are known plant defense metabolites (10, 46–48). Thus, our efforts were focused on the crystallization of TuUGT202A2 with various flavonoids to reveal the molecular basis underlying this enzyme's promiscuity and hence, its contribution to TSSM's broad range of plant hosts.

The first complex crystallized, UDP-(S)-naringenin (Fig. 4*A*) ($P2_12_12$, PDB code: 8GKN), was determined at 2.7 Å resolution. It contained two protein molecules in the asymmetric unit. In both chains, residues 11–437 could be modeled. Only residue 256 of chain B did not have a well-defined electron density. The overall conformation of TuUGT202A2 in both chains is almost identical, only differing in the region composed of residues 106–115. Electron densities for a molecule of UDP and (S)-naringenin were seen in each chain (Fig. S9). Note that the racemic mixture of naringenin was used for cocrystallization, but the (S) enantiomer fit the electron density better and was observed to form more favorable interactions with TuUGT202A2. Therefore, this enantiomer was included in the model. When comparing chains A and B to the apoform crystal structure, the RMSD calculated were 0.8 Å (over 406 superposed Cα atom pairs) and 0.9 Å (over 388 superposed Cα atom pairs), respectively. In the case of chain A, the biggest differences in main chain conformation correspond to the loop regions comprised of residues 76–82 and 247–262. On the other hand, chain B, in addition to these two loop regions, also differs in the conformation of fragment including residues 106–115. These structures suggest that the N-terminal and C-terminal domains of this TSSM's UGT do not need to undergo large conformational changes to be able to accommodate the activated sugar donor and acceptor in the catalytic site, in contrast to bacterial glycosyltransferases (42). Moreover, this process is most likely driven by the movement of certain loop regions that display high conformational flexibility.

Table 1
Summary of steady state kinetics for WT and R252A mutant using UDP-Glc and 1-naphthol

	UDP-Glc[a]		1-naphthol[a]	
Kinetic parameters	WT	R252A	WT	R252A
k_{cat} (s^{-1})	0.088 ± 0.023	0.028 ± 0.002	0.20 ± 0.02	0.040 ± 0.003
K_M (μM)	47 ± 13	94 ± 18	132 ± 16	100 ± 20
V_{max} (μM s^{-1})	(3.3 ± 0.3) × 10^{-3}	(1.1 ± 0.1) × 10^{-3}	(7.7 ± 0.5) × 10^{-3}	(1.5 ± 0.2) × 10^{-3}
k_{cat}/K_M (μM^{-1} s^{-1})	(1.9 ± 0.7) × 10^{-3}	(2.9 ± 0.6) × 10^{-4}	(1.5 ± 0.2) × 10^{-3}	(4.0 ± 0.9) × 10^{-4}

[a] Indicates the substrate whose concentration was varied.

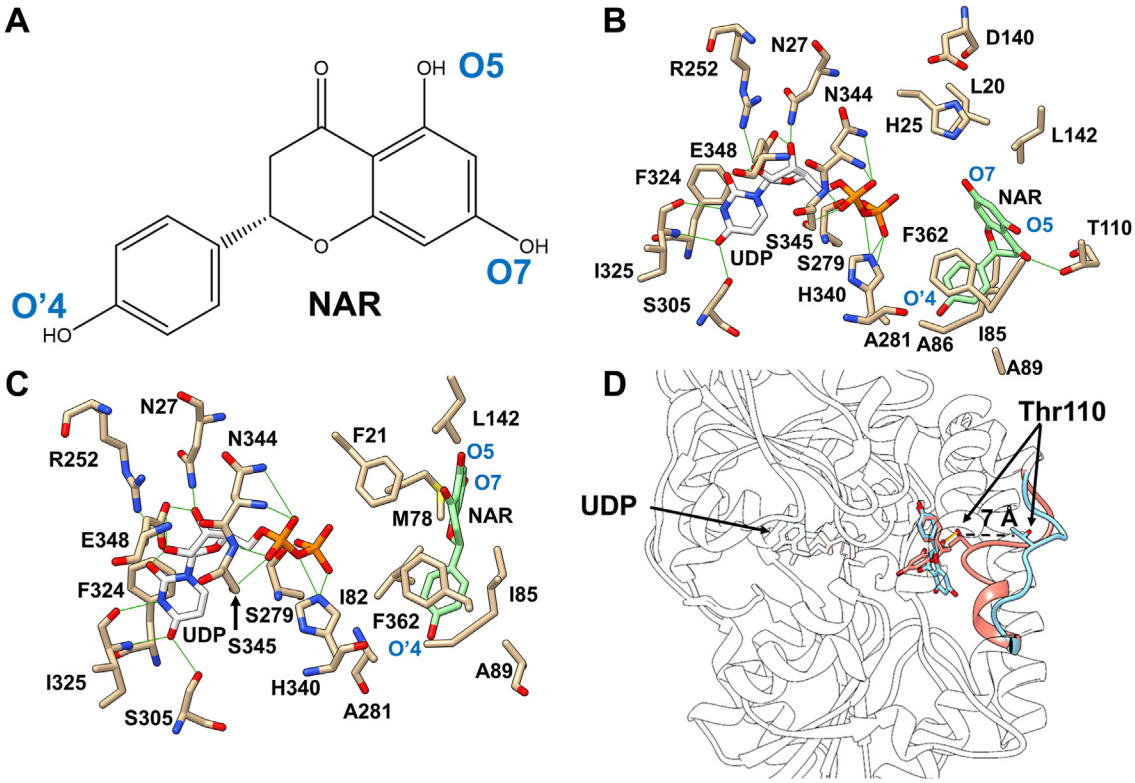

Figure 4. Structural analysis of the crystal complex of TuUGT202A2 in the presence of UDP and (S)-naringenin. *A*, chemical structure and labeling of hydroxyl groups of (S)-naringenin (NAR). *B*, chain A atomistic representation of the interactions between enzyme (carbon atoms in *tan color*) and the two ligands: UDP (carbon atoms in *white*) and NAR (carbon atoms in *green*). The observed hydrogen bonds between ligands and enzyme are shown in *green color*. Asp140 is displayed due to its importance in catalysis as part of the conserved His-Asp catalytic dyad, although it was not located within the cut-off distance. *C*, chain B atomistic representation of TuUGT202A2 catalytic site in the presence of UDP and NAR (coloring follows the same scheme as (*A*)). *D*, superimposition of chain A and B from PDB: 8GKN. Only (S)-naringenin and residues 106–115 have been colored to highlight the conformational difference between chain A, in *salmon*, and chain B, in *light blue*, and the different orientation of the ligand in each chain. The distances between the backbone of Thr110 from both chains were measured using UCSF Chimera. The predicted hydrogen bond between Thr110.A and (S)-naringenin is shown in a *solid yellow line*. UGT, UDP glycosyltransferase.

In this crystal structure, UDP is bound in both chains in a very similar orientation, consistent with the mode of binding seen in the UDP-Glc complex (PDB code: 8SFY). Compared to the UDP-Glc crystal structure, in this complex, besides the already mentioned Arg252, only one additional residue, Asn27, is observed to interact with UDP in both chains (Fig. 4, *B* and *C*). This interaction arises from the slight movement of its side chain. On the other hand, (S)-naringenin adopts two different orientations in each chain, suggesting that the catalytic pocket where the acceptor binds is highly flexible (Fig. 4, *B* and *C*). In chain A, residues Leu20, His25, Ile85, Ala86, Ala89, T110, Leu142, Ala281, and Phe362 form the binding site (using a 4 Å distance cut-off). Among these residues, only two directly interact with (S)-naringenin, Thr110 through a hydrogen bond with the keto group of the ligand and Phe362 through π- π interactions (Fig. 4*B*). In the case of chain B, residues Phe21, Met78, Ile82, Ile85, Ala89, Leu142, Ala281, and Phe362 form the hydrophobic pocket (using a 4 Å distance cut-off) (Fig. 4*C*). In this orientation, Phe362 still interacts with (S)-naringenin through π- π stacking, but no hydrogen bonding with any residue is observed.

Although this crystal structure contains the product of the reaction, UDP, instead of the substrate, UDP-Glc, we can explore whether any of these two chains represent the putative catalytically productive complex. To do this, we superimposed the structure of TuUGT202A2-UDP-Glc complex with each chain and measured the distance between the anomeric carbon and the acceptor's closest hydroxyl group. In chain A, the closest hydroxyl group is O7 and it is ∼4.6 Å away from the anomeric carbon. In addition, it is located 4 Å off His25, a residue that is part of a catalytic conserved dyad in GT-B enzymes that acts as the catalytic base to deprotonate the hydroxyl group (41, 49). Note that Asp140, the other residue of this catalytic dyad, is observed to form a hydrogen bond with His25, as expected during catalysis (Fig. 4*B*). In chain B, O5 is the closest hydroxyl group, but it is located 9 Å from the anomeric carbon and 7 Å from His25 (Fig. 4*C*). While the distances measured in chain B are too large for catalysis to take place, the distances measured in chain A are comparable with the ones reported in the active site of other GT-B catalytic complex crystal structures (41, 49, 50).

We investigated further whether the different modes of binding observed for (S)-naringenin in the crystal structure are of biological relevance and whether chain A indeed resembles a putative catalytically productive complex. Since a high degree of plasticity in the catalytic site usually leads to low regioselectivity, we studied the products of the reaction between TuUGT202A2 and naringenin using LC-MS. Compared to the

control reaction (no enzyme present), the chromatogram revealed two peaks with the corresponding mass of a molecule of glucose added and one peak with the corresponding mass for a double glycosylated product (Fig. S10). Naringenin contains three hydroxyl groups, meaning three possible places where TuUGT202A2 can catalyze the addition of a sugar moiety. Our LC-MS results indicated that at least two monoglycosylated isomers are being formed (we cannot rule out the possibility that a third isomer is co-eluting with one of these two). Commercially available naringenin-7-O-glucoside and naringenin-4′-O-glucoside were co-injected with the reaction mixture to determine the identity of the two monoglycosylated products being formed. Notwithstanding both isomers eluted very closely (retention times: 4.67 and 4.71 min), this experiment allowed us to conclude that naringenin-4′-O-glucoside is the first isomer eluting (retention time 4.67 min) and naringenin-7-O-glucoside (retention time: 4.71 min) is the second one (Fig. S11).

By combining the structural data with the functional experiments, we conclude that chain A of UDP-(S)-naringenin complex (PDB code: 8GKN) may indeed represent a putative catalytically productive complex in which (S)-naringenin will be modified at position O7. On the other hand, chain B represents an acceptor mode of binding that is not catalytically productive. As previously mentioned, at the protein level, the only difference between both chains is the conformation adopted by the residues 106–115. This loop seems to adopt a "closed" conformation in chain A and an "open" conformation in chain B. In this "closed" conformation, Thr110 forms a hydrogen bond with (S)-naringenin, whereas in the "open" conformation, this residue is located 7 Å further away (Fig. 4*D*). It is important to note that the conformation of the loop in chain B is not stabilized by any crystal contacts. While it is tempting to infer that chain B represents an intermediate stage in catalysis, where the acceptor is entering/leaving the active site, this orientation could also just be a result of crystallization and the highly plastic catalytic cavity of this enzyme.

Additional studies to deepen the understanding of TuUGT202A2 promiscuity were performed with two other flavonoids, kaempferol and quercetin, that were also part of the first *in vitro* substrate screening in which TuUGT202A2's activity was tested (10). LC-MS revealed that TuUGT202A2 can produce almost identical quantities of at least three different monoglycosylated isomers of kaempferol and one diglycosylated product (Fig. S12). On the other hand, TuUGT202A2 can add a molecule of glucose to quercetin in at least three different hydroxyl positions but displays a higher preference for a certain hydroxyl group *versus* the other two (Fig. S13). Interestingly, no double glycosylation products could be seen with quercetin as a substrate. This data further supports the idea that TuUGT202A2 has a very highly plastic catalytic cavity underlying its promiscuity.

Structural and functional studies of TuUGT202A2 interactions with naringin

The fact that TuUGT202A2 does not only have a low regioselectivity towards naringenin but also can double glycosylate this compound suggests that this enzyme can accommodate larger compounds in its active site, such as already decorated compounds. This is of biological relevance since many flavonoids exist as glycosyl-adducts in plants (51). Our working hypothesis stated that this enzyme's catalytic site is open-ended, allowing it to accommodate larger acceptors, such as flavonoids with extended glycan modifications. To test this hypothesis, we examined the activity of TuUGT202A2 towards two different types of naringenin disaccharides' derivatives: naringin, a disaccharide derivative with a 2-O-(alpha-L-rhamnopyranosyl)-beta-D-glucopyranosyl moiety at position 7 (O7) (Fig. 5*A*), and narirutin, a disaccharide derivative with a 6-O-(6-deoxy-alpha-L-mannopyranosyl)-beta-D-glucopyranosyl moiety at O7. LC-MS analysis revealed that both compounds are TuUGT202A2 substrates (Figs. S14 and S15). However, we were not able to determine which of the hydroxy groups (O5 or O'4) were modified.

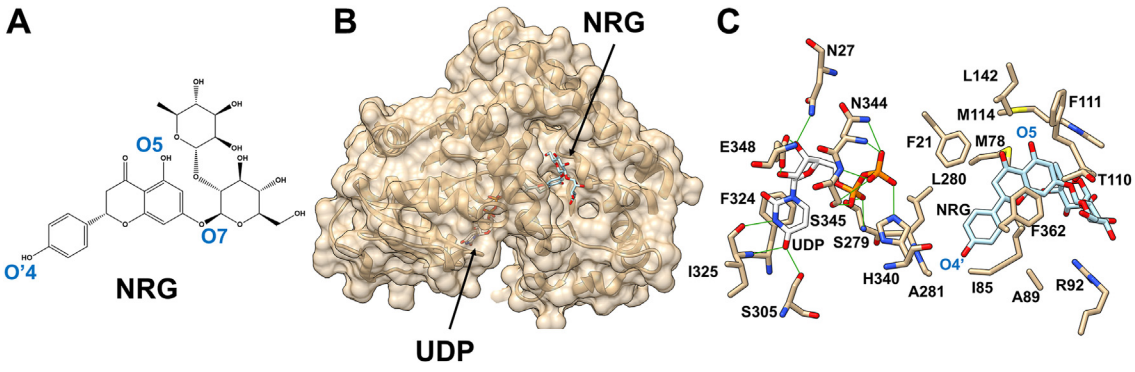

Figure 5. The crystal structure of TuUGT202A2 complex with UDP and naringin reveals the molecular basis of the open-ended catalytic site. Figures were made using only chain B of this crystal structure for simplicity, as chain A and B displayed identical conformation at the protein level and orientations of UDP and naringin in the catalytic site. *A*, chemical structure and hydroxyl labeling for naringin (NRG). *B*, surface representation of the crystal structure of TuUGT202A2 in complex with NRG (carbon atoms in *blue*) and UDP (carbon atoms in *white*). The open-ended catalytic site allows to accommodate naringenin backbone while leaving the glycan portion solvent exposed. *C*, atomistic representation of UDP and NRG mode of binding. NRG-binding pocket includes some unique residues that were not part of any of the chains of UDP-(S)-naringenin crystal structure. This further proves the plasticity of this enzyme's catalytic site. Carbon atoms for UDP are colored *white* and NRG are in *light blue color*. Observed hydrogen bonds are colored *green*. UGT, uridine glycosyltransferase.

To elucidate the molecular basis of this open-ended catalytic site, co-crystallization experiments were performed with both substrates in the presence of UDP. The crystal structure of UDP-naringin complex was determined at 1.8 Å resolution ($P2_12_12$, PDB code: 8SFU). The structure contained two molecules in the asymmetric unit. The interpretable electron density for the enzyme's backbone starts at residue 11 and ends at residue 437 in both chains. Residues 250–257 lack electron density in chain A, likewise residues 251–257 in chain B. Within each chain, electron densities corresponding to a molecule of UDP and naringin could be easily distinguished (Fig. S16). The overall conformation of TuUGT202A2 seen in both chains is nearly identical, as well as the orientations of UDP and naringin among these two chains. This crystal structure reveals the nature of the open-ended catalytic site, as it can clearly be seen how the naringenin portion is enclosed in the active site, while the disaccharides are solvent exposed (Fig. 5*B*). Interestingly, the regions of the enzyme that underly this mode of binding are the ones belonging to the N-terminal end domain that are not part of the conserved Rossmann-like fold and distinguished this spider mite UGT from bacterial UGTs. There are two main differences when comparing the binding of UDP in this crystal structure to the one described for UDP-(S)-naringenin complex. First, it is missing the hydrogen bond between Arg252 and the ribose since this residue could not be modeled (Fig. 5*C*). Secondly, the positioning of the diphosphate group is different between these two structures. For instance, in the UDP-naringin complex, the β-phosphate occupies the same position as the α-phosphate in the UDP-(S)-naringenin structure. This movement of the diphosphate group is accompanied by a shift in the conformation of the Ser279 side chain that now interacts with the α-phosphate, instead of the β-phosphate (Fig. 5*C*). From a catalytic point of view, this positioning of UDP would locate the glucose group further away from the acceptor. While it is tempting to suggest that UDP was trapped in this crystal structure in a state that resembles the activated sugar donor entering/leaving the active site and that is why the entire flexible loop composed by residues 247–262 could not be completely modeled, it cannot be ruled out that this is just a product of crystallization, as this region is near crystal contacts.

The structure of the complex with naringin not only revealed the open-ended catalytic site but also supports the idea that this enzyme holds a very plastic catalytic site. The binding pocket is composed of residues Phe21, Met78, Ile85, Ala89, Arg92, Thr110, Phe111, Met114, Leu142, Ala281, and Phe362 (using a 4 Å distance cut-off) (Fig. 5*C*). Among these residues, direct interactions take place between the backbone of Thr110 and the ligand through a H-bond, as well as Phe362 and naringin π-π stacking. Arg92, Phe111, and Met114 were not part of the binding pocket in either of the chains in the UDP-(S)-naringenin structure. In this crystal complex, the distance between the closest hydroxyl group of naringin, 4′O, and the putative location of the anomeric carbon of UDP-Glc is 9 Å, while this hydroxyl group is located 12 Å from the catalytic His25. These distances are too far for catalysis to occur and thus, this crystal structure does not resemble a catalytically productive complex. However, our functional studies demonstrated that naringin is indeed a substrate of this enzyme and it would make sense that the enzyme catalyzes the reaction in the 4′O, which is less hindered, and we have demonstrated that it can catalyze this reaction (see section: Structural and functional studies of TuUGT202A2 interactions with naringenin). We suggest that the mode of binding of naringin highlights the flexibility of the enzyme's active site and its ability to accommodate substrates in many different orientations.

Discussion

Arthropods represent a large and diverse group of species that significantly impacts human health and agriculture. While the role of human UGTs in phase II detoxification has been extensively studied (5, 11–13), their contribution to arthropod xenobiotic resistance is less known. Spider mites are one of the most polyphagous pests, as they can feed on more than 1100 plant species (19). Short life cycle, high fecundity, haplodiploid sex determination, and a large metabolic detoxifying machinery are the main biological properties responsible for their extraordinary adaptation potential and devastating effect on crops. This pest developed resistance to virtually all chemical classes used for its control (18, 52). Availability of the whole genome sequences enabled the characterization of the detoxification inventory in many subsequent studies (reviewed in (53)), which is enriched by horizontal gene transfer, such as glycosyltransferases (10). It is important to note that the majority of studies, that investigate the detoxifying machinery of *T. urticae*, focus on the gene level and there is currently a lack of knowledge on the molecular details of the enzymes that contribute to this xenobiotic metabolism process (54).

In this work, we were able to elucidate the first structure of a UGT originating from *T. urticae*. TuUGT202A2 displayed the characteristic GT-B folding, consisting of two β/α/β Rossmann-like fold domains facing each other and creating the catalytic site in between them. It is important to point out that while the conserved Rossmann-like fold present in both the N- and C-terminal domains were very similar between TuUGT202A2 and bacterial glycosyltransferases, major differences were seen in regions that do not form part of this conserved fold, located within the N-terminal domain and underly the open-ended nature of the catalytic site. These differences are believed to be a product of evolution and clearly distinguish bacterial UGTs from TuUGT202A2. The N-terminal domain of UGTs is believed to be responsible for substrate specificity and to this date, an experimental model of this domain originating from an animal has not been experimentally elucidated. Our structural analysis using AlphaFold2 theoretical models of full-length human UGTs suggests that the N-terminal domain of this arthropod UGT may be similar to the human ones and potentially other animals. While this cannot be proven until an experimental structure of the N-terminal domain of another animal is elucidated, TuUGT202A2 may serve as a model for the

organization and structural flexibility of the animal UGTs' acceptor-binding site.

One of the biggest challenges in the analysis of glycosyltransferase is their functional characterization. Currently, acceptor and donor specificity for more than 95% of glycosyltransferases are not known (6). This is further aggravated by a small number of complex structures that contain a sugar donor and an acceptor. In this work, we were able to elucidate the structure of TuUGT202A2 in different catalytically relevant stages. First, the crystal structure of the apoform was elucidated. This structure demonstrated that the hydrophilic pocket where UDP-Glc binds is readily open. Secondly, the crystal complex containing UDP-Glc was obtained. Structural, functional, and computational studies support the idea that upon UDP-Glc binding to this readily open cavity, a loop (residues 249–262) undergoes major conformational changes for catalysis to take place. The displacement of this loop allows Arg252 to hydrogen bond with the ribose part of UDP-Glc. Kinetic studies using an Arg252Ala mutant suggest that this interaction is needed not only for proper binding of UDP-Glc and activation but also to maintain this loop in the "closed" conformation, in which the catalytic complex is protected. Arg252 is conserved among all spider mite UGTs, which suggests that this mechanism extrapolates to all these UGTs (Fig. S7). To our knowledge, this work is the first one to reveal this lid-like mechanism in the GT-B superfamily. However, this mechanism cannot be concluded to be unique to spider mite UGTs. Many PDB deposits of UGTs from other organisms lack partial or full electron density for this region which makes it very challenging to determine whether other species use a similar mechanism; plus most studies focus on the N-terminal end domain and the mechanism of binding of the acceptor compound and overlook conformational changes in the C-terminal domain.

Furthermore, two additional TuUGT202A2 crystal complexes were obtained, one included UDP and (S)-naringenin and the other included UDP and naringin. Pairing structural analysis with the identification of the reaction products enabled us to conclude that chain A of the UDP-(S)-naringenin crystal structure may represent a catalytically productive complex. The rest of the crystal complexes, chain B of UDP-(S)-naringenin and UDP-naringin, are believed to be products of the high plasticity of the acceptor-binding site and most likely they do not mimic productive complexes. This enzyme displays a promiscuity characteristic of "drug-metabolizing" enzymes (55). Other UGTs, from human, plants, and bacteria, have also been shown to possess this property (34, 41, 56). Notably, we successfully elucidated the molecular basis of TuUGT202A2 promiscuity by determining the flexibility and open-ended nature of its catalytic site, which allows to accommodate a wide range of substrates in different orientations, including already decorated ones. Working with sugar derivatives of flavonoids is particularly challenging for two main reasons: first, as our LC chromatograms show that they tend to elute very close to each other. In fact, the idea that some of these isomers are co-eluting resulting in a single peak cannot be ruled out, which will result in an even lower regioselectivity than what our results suggest. Secondly, not all isomers are commercially available and use of some that are available can be cost prohibited. While one limitation of this study is that the nature of all products could not be identified, this restriction does not take away from the main idea of this manuscript: a plastic and open-ended catalytic site underlies the substrate promiscuity of TuUGT202A2, which was proved by structural and functional studies.

In summary, this paper presents the first crystal structure of a UGT originating from a chelicerate. The structural analysis performed pointed out important structural differences in the acceptor-binding domain when compared to bacterial homologs. Interestingly, these features may be characteristic of animal UGTs, as they are also present in theoretical models of human UGTs. The structural and functional characterization of this enzyme enabled us to determine the molecular basis of substrate promiscuity, as its broad, highly flexible, and open-ended acceptor-binding site allows for the incorporation of a wide range of compounds in different orientations. Furthermore, the crystallization of this enzyme, not only in its apoform but also in other catalytically relevant stages, revealed important conformational changes that this enzyme undergoes for catalysis to occur. This work does not only represent a steppingstone on the understanding of host-shift and acaricide adaptation of a major agricultural pest but due to the structure similarity among UGTs across diverse species, and the fact that we are the first to reveal the folding of the N-terminal end of a UGT originated from an animal, the impact of this work reaches far beyond the understanding of TSSM.

Experimental procedures

Protein production and purification

All reagents were purchased from Sigma-Aldrich, Santa Cruz Biotechnology, or Thermo Fisher Scientific. The pET100 Directional TOPO (pET100-D-Topo) plasmid containing Tetur22g00270 (UniProt: T1KUK4) was transformed into *E. coli* strain BL21(DE3) by heat shock method and spread on 2% agar plates (100 μg/ml ampicillin). A single colony of BL21(DE3) cells harboring pET100-D-Topo-tetur22g00270 was picked up and grown overnight in 5 ml LB medium (100 μg/ml ampicillin) supplemented with 50 mg of glucose. Then, 1 l of terrific broth media supplemented with 100 μg/ml ampicillin was inoculated with 5 ml of overnight cultures and incubated at 37 °C, 200 rpm, until the A_{600} reached approximately 0.8. Cultures were cooled to 16 °C and protein expression was induced with 0.4 mM IPTG. Protein production was continued overnight at 16 °C with shaking at 150 rpm. Cultures were centrifuged at 13,000*g* for 15 min at 4 °C. Supernatants were discarded, while pellets were collected and frozen at −80 °C until further use.

Bacterial cell pellets were thawed and resuspended in 5 ml of lysis buffer (20 mM Na_2HPO_4, 500 mM NaCl, 10 mM imidazole, 2% (v/v) glycerol, pH 7.9) per 1 g of the pellet. Solutions were lysed on ice by sonication using a Q500 Sonicator (Qsonica) for ten cycles. Each cycle had 10 s of sonication and 50 s rest on 250 W. The cell debris was removed by

centrifugation (36,000*g*, 40 min, 4 °C). TuUGT202A2 construct was designed to contain a poly-histidine tag at the N-terminal end to help with purification. Thus, the supernatants were run through HisPur Ni-NTA Resin (Thermo Fischer Scientific) in Econo-Pac Chromatography Columns (Bio-Rad) which had previously been equilibrated with lysis buffer. The Ni-NTA resin was washed with wash buffer (20 mM Na_2HPO_4, 500 mM NaCl, 30 mM imidazole, 2% (v/v) glycerol, pH 7.9) to remove bound contaminants, and (His)6-UGT was eluted from the column with elution buffer (20 mM Na_2HPO_4, 500 mM NaCl, 250 mM imidazole, 2% (v/v) glycerol, pH 7.9). Protein concentration was estimated with the Pierce Coomassie Plus (Bradford) Assay Reagent (Thermo Fisher Scientific), whereas protein content and purity of samples were inspected with 12% SDS-PAGE. The samples containing target protein were further purified on HiLoad 16/600 Superdex 200 pg size-exclusion column attached to an ÄKTA Pure protein purification system (GE Healthcare) using 20 mM Na_2HPO_4, 150 mM NaCl, pH 7.9. Fractions containing UGT were pooled together, concentrated using Amicon Ultra centrifugal filter units with 10 kDa MWCO (Millipore), and sample purity was estimated with 12% SDS-PAGE (Fig. S17). Protein concentration was determined using the molecular weight and extinction coefficient estimated by ProtParam (57) with NanoDrop 2000c Spectrophotometer (Thermo Fisher Scientific).

The primers and annealing temperatures used for site-directed mutagenesis are shown in Table S4. The PCR was performed using a PTC-100 PCR Programmable Thermal Controller in a Hard Shell 96-Well PCR plate (Bio-Rad) sealed with Bio-Rad Microseal PCR Plate Sealing Film. The protocol followed was similar to the previously described one (58). The PCR product was analyzed using 1% agarose gel. After confirming the correct size of the amplification product, the PCR products were methylated and ligated and later transformed to DH5α, as previously described (58). Plasmids from single colonies were purified and the presence or absence of the mutation was confirmed by sequencing the construct using T7 promoter and T7-terminator using the plasmid sequencing services from Eton Bioscience or Genewiz. Purified DNA confirmed to have the mutation was transformed into BL21(DE3), as mentioned above. The mutant was expressed and purified following the same procedure as the WT (Fig. S18).

Crystallization

Preliminary crystallization conditions, for the apoform protein, were tested by sitting drop vapor diffusion method and commercially available screens: Index, Natrix, PEG/Ion (Hampton Research), PACT premier (Molecular Dimensions), Wizard Cryo, Wizard Classic, and JCSG+ (Rigaku Reagents). Crystallization drops were set up in protein:mother liquor ratio 1:1 at 4 °C or 20 °C on Intelli-Plate 96-2 Original crystallization plates. The apoform of TuUGT202A2 was finally obtained in 5% Tacsimate, pH 7.0, 0.1 M Hepes, 10% w/v PEG 5000. Different techniques were used to trap the different substrates in the crystal structure. TuUGT202A2 and UDP-Glc complex was obtained by adding UDP-α-D-Glucose powder (Cayman Chemicals) to the well containing native protein crystals. TuUGT202A2, UDP, and naringin/naringenin were obtained following the same procedure as with apoform protein, but the enzyme was incubated with 1 mM UDP (Cayman Chemicals) and 2.5 mM naringin (Sigma-Aldrich) or 5 mM (±)-naringenin (Sigma-Aldrich), for 30 min at 4 °C prior to crystallization. The conditions that yielded the crystal structures were as follows: 0.2 M NaCl, 0.1 M Bis-tris pH 6.5, 25% w/v PEG 3350; and, 0.2 M lithium sulfate monohydrate, 0.1 M Bis-tris pH 6.5, and 25% w/v PEG 3350, for UDP-naringin and UDP-naringenin, respectively. No cryoprotectant was used in any case. All crystals were flash-cooled in liquid nitrogen.

Data collection, structure determination, and model refinement

Table S5 shows the data collections statistics for determined crystal structures of TuUGT202A2 and complexes structures. Crystals were cryo-cooled in liquid nitrogen and most data was collected using the Southeast Regional Collaborative Access Team 22ID beamline, and Structural Biology Center (SBC-CAT) at Argonne National Laboratory 19BM beamline was used in the case of TuUGT202A2 and UDP-Glc complex. Datasets were processed with the HKL-3000 software package (59). Molecular replacement for 6PNT was performed using MOLREP (60) integrated with HKL-3000 (61) and the PDB entry 2IYA as a starting model. BUCCANEER (62) and HKL-3000 were used to rebuild the initial model. Refinement was performed using REFMAC (63) and HKL-3000. When applicable, the noncrystallographic symmetry was used during the whole process of refinement. TLS refinement was used during the last stages of refinement and the TLS Motion Determination server was used for partitioning protein chains into the rigid bodies undergoing vibrational motions (64). The model was updated and validated with COOT (65). MOLPROBITY was used in the final steps of the model validation (66). For all complexes containing flavonoids, a similar approach was used. In this case, all data sets were collected using SER-CAT 22ID beamline and processed using HKL-3000. PDB entry 6PNT was used as the initial model for molecular replacement for UDP-(S)-naringenin structure, and afterward, chain A of UDP-(S)-naringenin structure was used for the complex containing naringin. The crystal structures of TuUGT202A2 were deposited in the Protein Data Bank (PDB) (67) with the accession codes 6PNT, 8SFY, 8GKN, and 8SFU.

Computational analysis of TuUGT202A2 crystal structures

COOT, PyMOL (68), and UCSF-Chimera (69) were used to visualize and analyze the structures, as well as to generate figures. DALI and PDBeFold were used to search for similar structures. Jalview (70) was used to elaborate the MSA between homologous sequences, whereas ConSurf (71) was used for conservation analysis between all UGTs originating from spider mite, whose full sequence was available in UniProt database (72). AlphaFold2 model of human UGTs were

obtained from UniProt as well and only a few residues were modeled with a pLDDT score lower than 50. The pI values were calculated using the ExPASy ProtParam tool. PDBePISA (73) was used to analyze the crystal structures and evaluate the formation of quaternary structures in solution. FoXS and MultiFOXS (40, 74) were used to analyze small angle X-ray scattering data. ModWeb was used to generate a model of the full construct (75). LigPlot was used to generate 2D representations of the ligand-binding site (76).

Size-exclusion chromatography-small angle X-ray scattering

SAXS was performed at BioCAT (beamline 18ID at the Advanced Photon Source, Chicago) with in-line SEC-SAXS to separate the sample from aggregates and other contaminants thus ensuring optimal sample quality. Five hundred microliters of the sample at a concentration of 5 mg/ml were loaded onto a Superdex 200 Increase 10/300 GL column (Cytiva), which was run at 0.6 ml/min by an AKTA Pure FPLC (GE) and the eluate after it passed through the UV monitor was flown through the SAXS flow cell. The flow cell consists of a 1.0 mm ID quartz capillary with ∼20 μm walls. A coflowing buffer sheath is used to separate sample from the capillary walls, helping prevent radiation damage (77). Scattering intensity was recorded using an Eiger2 XE 9M (Dectris) detector which was placed 3.654 m from the sample giving us access to a q-range of 0.0027 $Å^{-1}$ to 0.42 $Å^{-1}$. 0.5 s exposures were acquired every 1 s during elution and data was reduced using BioXTAS RAW 2.1.1 (78). Buffer blanks were created by averaging regions flanking the elution peak and subtracted from exposures selected from the elution peak to create the I(q) *versus* q curves used for subsequent analyses. Table S6 includes additional information regarding the data collection. Evolving factor analysis was used for deconvolution of the data and MultiFoXS to generate theoretical models that best represent the experimental data.

Activity assay

1-naphthol was the model substrate used to obtain kinetic parameters. The formation of the corresponding glucoside was measured by fluorescence spectrophotometry, using excitation/emission wavelength of 287/335 nm, in a 96-well plate with a Bio-Tek Synergy H1 multimode microplate reader, in a similar fashion previously described (10). These kinetic assays were done in pseudo-first order by keeping UDP-Glc or 1-naphtol at saturated conditions (2 mM) and varying the concentration of the other substrate from 0 to 300 μM, in the case of 1-napthol and 0 to 600 μM in the case of UDP-Glc. The reaction was incubated at room temperature for 5 min. The total volume reaction was 250 μl and contained 0.5 μg of the enzyme. This reaction was performed in 20 mM sodium phosphate buffer, 150 mM NaCl, pH 7.8. 1-naphthyl-β-D-glucoside sodium salt was purchased from Santa Cruz Biotechnology, Inc and used to create a standard curve to relate relative fluorescence units to quantities of product being formed. Data was fit to the Michaelis–Menten model for all enzyme kinetics studies using OriginPro software (https://www.originlab.com/). The Michaelis–Menten graphs are included in Fig. S19, and no cooperativity was observed when data was fit to the Hill equation.

For the analysis of the products of the reactions, a reaction mixture containing 25 μg of the enzyme, 1 mM UDP-Glc, and 1 mM acceptor substrate was incubated at room temperature for an hour. Afterward, the reaction was stopped by adding three volumes of 50:50 water:methanol. The mixture was centrifuged for 15 min and diluted by adding more 50:50 water:methanol to a final concentration of 5 μM prior to LC-MS analysis on with Waters Xevo G2-XS QTof with C18 column. Ten microliters of the mixture was injected and run at a flow rate of 0.3 ml/min using water with 0.1% formic acid as solvent A and acetonitrile as solvent B. The gradient started at 90%A-10%B, increased to 50%A-50%B in a span of 5 min, held at that gradient for 2 min, and then ramped up to 100%B for 1 min to then return to the initial gradient. MS was set to detect a range of masses from 50 to 1500 Da and the energy of collision used in MS was a ramp of 10 to 60 V, with a scanning time of 0.2 s. For all experiments, a control reaction (nonenzymatic) was performed following the same procedure. Co-injection experiments were conducted in a similar fashion, but the reaction mixture was spiked with the standard: naringenin-4′O-glucoside (Extrasynthese) or naringenin-7-O-glucoside (Gifted from Julieta Petrich, Grotewold lab at MSU).

Isothermal titration calorimetry

All calorimetric experiments were conducted on an affinity isothermal titration calorimeter from TA Instruments. Before titration, the reference cell was filled with double-distilled water. All solutions were filtered and degassed for 10 min under vacuum. The sample cell was filled with 350 μl of 80 μM enzyme in 20 mM phosphate buffer pH 7.5, and the syringe was loaded with 5.0 mM UDP-Glc in 20 mM sodium phosphate pH 7.5. During the titration, the reaction mixture was continuously stirred at 150 rpm, and the injection volume was 2 μl for a total of 20 injections at room temperature. The background titration profiles were obtained by injecting 5.0 mM UDP-Glc into 20 mM sodium phosphate pH 7.5. The amount of heat produced per injection was calculated by integration of the area under individual peaks by the Launch NanoAnalyze software (https://www.tainstruments.com) provided with the instrument. Results are presented as averages from three independent experiments.

MD simulations

Prior to MD simulations, each protein model was placed in a virtual box with 10 Å margins between the protein and the box walls. Then, to mimic the physiological conditions, explicit TIP3P water molecules were added to the box. To neutralize the system charge and create ionic strength typical for physiological fluids, that is, 0.15 M, appropriate number of Na^+ and Cl^- ions were added. The protein was described with the ff14sb amber force field, whereas the substrates were parametrized within the GAFF2 force field. All of the steps described above were taken with the tleap program from the AmberTools

package (79). Thus, prepared models were subjected to energy minimization during three consecutive stages: first, just the water molecules and ions underwent relaxation (with atomic positions of the protein complex restrained with a force constant of 500 kcal/molÅ^2), then protein complexes were relaxed, first with the force constant of positional restraints of 10 kcal/molÅ^2, and then with no restrains applied. After energy minimization, heating of the system from 0 to 300 K was simulated during a 100-ps-long MD simulation under constant volume conditions with a Langevin thermostat and the system then underwent a second, 1-ns-long equilibration at a constant pressure of 1 bar. During the heating and density equilibration, the backbone of the protein was restrained with a force constant of 1 kcal/molÅ^2. Then, the production stage of the MD simulations followed (102 ns). During this stage of the MD, pressure and temperature were kept constant at 1 bar and 300 K, respectively, the frequency of saving current geometry to a data file was set to once per every 5000 steps, and the SHAKE algorithm was used to constrain values of valence bonds and angles including hydrogen atoms. The time step of the simulation was set to 2 fs, long range electrostatic was calculated with the PME method, whereas for direct nonbonded interactions, the default cut-off radius of 8 Å was used. Each simulation was repeated three times with a different seed of pseudo random number generator. The GPU version of pmemd program from the AMBER package was used (80, 81). To analyze the MD trajectories, the cpptraj program from AmberTools was used, particularly to (1) cluster the MD geometries according to structural similarities using the HierAgglo algorithm and (2) to calculate the RMSD values for the protein backbone.

Data availability

Data will be made available on request.

Supporting information—This article contains supporting information (78, 82–88).

Acknowledgments—We would like to thank Dr Caryn Outten (University of South Carolina) for giving access to the microplate reader, Dr Grotewold lab at MSU (especially visiting scholar Julieta Petrich) for gifting us some of the flavonoids glycosylated derivatives to use as standards, the Michigan State University mass spectrometry center, especially Dr Anthony Schilmiller, for his help and guidance in the reporting the mass-spectrometry data, as well as Dr Bill Cotham from the University of South Carolina mass spectrometry center.
This project was funded by USDA's National Institute of Food and Agriculture, award # 2020-67014-31179 through the NSF/NIFA Plant Biotic Interactions Program.
Structural results shown in this report are derived from data collected at Southeast Regional Collaborative Access Team (SER-CAT; 22 ID) or Structural Biology Center (SBC-CAT) at Argonne National Laboratory 19-BM. Supporting institutions may be found at www.ser-cat.org/members.html. Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract Nos. DE-AC02-06CH11357 and W-31-109-Eng-38.
This research used resources of the Advanced Photon Source; a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. This project was supported by grant P30 GM138395 from the National Institute of General Medical Sciences of the National Institutes of Health.
This work was partially supported by ASPIRE II and ASPIRE III grants from the Office of the Vice President of Research at the University of South Carolina.
This research was supported in part by PL-Grid Infrastructure. Computations were performed at Academic Computer Centre Cyfronet AGH. The content is solely the responsibility of the authors and does not necessarily reflect the official views of the National Institute of General Medical Sciences or the National Institutes of Health.

Author contributions—R. H. A. and M. C. conceptualization; R. H. A., B. A., M. P., L. D., M. B. W., J. B. H., and T. B. investigation; R. H. A., B. A., M. P., L. D., M. B. W., J. B. H., and T. B. methodology; R. H. A., M. B. W., J. B. H., T. B., and M. C. formal analysis; R. H. A., M. B. W., J. B. H., T. B., and M. C. visualization; A. K., S. S., T. V. L., M. G., V. G., and T. B. resources. R. H. A., T. B., and M. C. writing–original draft; B. A., M. P., L. D., A. K., S. S., M. B. W., J. B. H., T. V. L., M. G., V. G., T. B., and M. C. writing–review and editing; M. G., V. G., and M. C. funding acquisition; M. C. supervision.

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

Abbreviations—The abbreviations used are: GT, glycosyltransferase; MD, molecular dynamics; MSA, multiple sequence alignment; SEC-SAXS, size-exclusion chromatography-small angle X-ray scattering; TSSM, two-spotted spider mite; UDP-Glc, UDP-Glucose; UGT, UDP glycosyltransferase.

References

1. Gu, X., Chen, L., Wang, X., Liu, X., You, Q., Xi, W., *et al.* (2014) Direct glycosylation of bioactive small molecules with glycosyl iodide and strained olefin as acid scavenger. *J. Org. Chem.* **79**, 1100–1110
2. Desmet, T., Soetaert, W., Bojarova, P., Kren, V., Dijkhuizen, L., Eastwick-Filed, V., *et al.* (2012) Enzymatic glycosylation of small molecules: challenging substrates require tailored catalysts chemistry. *Chem. Eur. J.* **18**, 10786–10801
3. Lim, E. K., and Bowles, D. J. (2004) A class of plant glycosyltransferases involved in cellular homeostasis. *EMBO J.* **23**, 2915–2922
4. Dianna, B., Isayenkova, J., Lim, E. K., and Brigitte, P. (2005) Glycosyltransferases: managers of small molecules. *Curr. Opin. Plant Biol.* **8**, 254–263
5. Meech, R., Hu, D. G., McKinnon, R. A., Mubarokah, S. N., Haines, A. Z., Nair, P. C., *et al.* (2019) The UDP-glycosyltransferase (UGT) superfamily: new members, new functions, and novel paradigms. *Physiol. Rev.* **99**, 1153–1222
6. Lairson, L. L., Henrissat, B., Davies, G. J., and Withers, S. G. (2008) Glycosyltransferases: structures, functions, and mechanisms. *Annu. Rev. Biochem.* **77**, 521–555
7. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrissat, B. (2014) The carbohydrate-active enzymes database (CAZy). *Nucleic Acids Res.* **42**, D490–D495
8. Messner, B., Thulke, O., and Schäffner, A. R. (2003) Arabidopsis glucosyltransferases with activities toward both endogenous and xenobiotic substrates. *Planta* **217**, 138–146
9. Nagare, M., Ayachit, M., Agnihotri, A., Schwab, W., and Joshi, R. (2021) Glycosyltransferases: the multifaceted enzymatic regulator in insects. *Insect Mol. Biol.* **30**, 123–137

10. Snoeck, S., Pavlidi, N., Pipini, D., Vontas, J., Dermauw, W., and Van Leeuwen, T. (2019) Substrate specificity and promiscuity of horizontally transferred UDP-glycosyltransferases in the generalist herbivore Tetranychus urticae. *Insect Biochem. Mol. Biol.* **109**, 116–127
11. Miley, M. J., Zielinska, A. K., Keenan, J. E., Bratton, S. M., Radominska-Pandya, A., and Redinbo, M. R. (2007) Crystal structure of the cofactor-binding domain of the human phase II drug-metabolism enzyme UDP-glucuronosyltransferase 2B7. *J. Mol. Biol.* **369**, 498–511
12. Jancova, P., Anzenbacher, P., and Anzenbacherova, E. (2010) Phase II drug metabolizing enzymes. *Biomed. Pap. Med. Fac. Univ. Palacky Olomouc Czech Repub.* **154**, 103–116
13. Yang, N., Sun, R., Liao, X., Aa, J., and Wang, G. (2017) UDP-glucuronosyltransferases (UGTs) and their related metabolic cross-talk with internal homeostasis: a systematic review of UGT isoforms for precision medicine. *Pharmacol. Res.* **121**, 169–183
14. Yonekura-Sakakibara, K., and Hanada, K. (2011) An evolutionary view of functional diversity in family 1 glycosyltransferases. *Plant J.* **66**, 182–193
15. Srivastava, P., Garg, A., Misra, R. C., Chanotiya, C. S., and Ghosh, S. (2021) UGT86C11 is a novel plant UDP-glycosyltransferase involved in labdane diterpene biosynthesis. *J. Bio Chem.* **297**, 101045
16. Smith, A. D., Page, B. D. G., Collier, A. C., and Coughtrie, M. W. H. (2020) Homology modeling of human uridine-5′-diphosphate-glucuronosyltransferase 1A6 reveals insights into factors influencing substrate and cosubstrate binding. *ACS Omega* **5**, 6872–6887
17. Breton, C., Šnajdrová, L., Jeanneau, C., Koča, J., and Imberty, A. (2006) Structures and mechanisms of glycosyltransferases. *Glycobiology* **16**, 29R–37R
18. Van Leeuwen, T., Tirry, L., Yamamoto, A., Nauen, R., and Dermauw, W. (2015) The economic importance of acaricides in the control of phytophagous mites and an update on recent acaricide mode of action research. *Pestic. Biochem. Physiol.* **121**, 12–21
19. Migeon, A., Nouguier, E., and Dorkeld, F. (2010) Spider mites web: a comprehensive database for the Tetranychidae. In: Sabelis, M. W., Bruin, J., eds. *Trends in Acarology*, Springer Netherlands, Dordrecht: 557–560
20. Dermauw, W., Wybouw, N., Rombauts, S., Menten, B., Vontas, J., Grbic, M., *et al.* (2013) A link between host plant adaptation and pesticide resistance in the polyphagous spider mite Tetranychus urticae. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 113–122
21. Grbić, M., Van Leeuwen, T., Clark, R. M., Rombauts, S., Rouzé, P., Grbić, V., *et al.* (2011) The genome of *Tetranychus urticae* reveals herbivorous pest adaptations. *Nature* **479**, 487–492
22. Ahn, S. J., Dermauw, W., Wybouw, N., Heckel, D. G., and Van Leeuwen, T. (2014) Bacterial origin of a diverse family of UDP-glycosyltransferase genes in the *Tetranychus urticae* genome. *Insect Biochem. Mol. Biol.* **50**, 43–57
23. Ji, M., Vandenhole, M., De Beer, B., De Rouck, S., Villacis-Perez, E., Feyereisen, R., *et al.* (2023) A nuclear receptor HR96-related gene underlies large trans-driven differences in detoxification gene expression in a generalist herbivore. *Nat. Commun.* **14**, 4990
24. Panche, A. N., Diwan, A. D., and Chandra, S. R. (2016) Flavonoids: an overview. *J. Nutr. Sci.* **5**, e47
25. War, A. R., Paulraj, M. G., Ahmad, T., Buhroo, A. A., Hussain, B., Ignacimuthu, S., *et al.* (2012) Mechanisms of plant defense against insect herbivores. *Plant Signal Behav.* **7**, 1306–1320
26. Treutter, D. (2006) Significance of flavonoids in plant resistance: a review. *Environ. Chem. Lett.* **4**, 147–157
27. Ullah, A., Munir, S., Badshah, S. L., Khan, N., Ghani, L., Poulson, B. G., *et al.* (2020) Important flavonoids and their role as a therapeutic agent. *Molecules* **25**, 5243
28. Dias, M. C., Pinto, D. C. G. A., and Silva, A. M. S. (2021) Plant flavonoids: chemical characteristics and biological activity. *Molecules* **26**, 5377
29. Ravishankar, D., Salamah, M., Akimbaev, A., Williams, H. F., Albadawi, D. A. I., Vaiyapuri, R., *et al.* (2018) Impact of specific functional groups in flavonoids on the modulation of platelet activation. *Sci. Rep.* **8**, 9528
30. Wright, B., Spencer, J. P. E., Lovegrove, J. A., and Gibbins, J. M. (2013) Insights into dietary flavonoids as molecular templates for the design of anti-platelet drugs. *Cardiovasc. Res.* **97**, 13–22
31. Wright, B., Moraes, L. A., Kemp, C. F., Mullen, W., Crozier, A., Lovegrove, J. A., *et al.* (2010) A structural basis for the inhibition of collagen-stimulated platelet function by quercetin and structurally related flavonoids. *Br. J. Pharmacol.* **159**, 1312–1325
32. Kawaii, S., Tomono, Y., Katase, E., Ogawa, K., and Yano, M. (1999) Antiproliferative activity of flavonoids on several cancer cell lines. *Biosci. Biotechnol. Biochem.* **63**, 896–899
33. Luque, T., Okano, K., and O'Reilly, D. R. (2002) Characterization of a novel silkworm (*Bombyx mori*) phenol UDP-glucosyltransferase. *Eur. J. Biochem.* **269**, 819–825
34. Hyung Ko, J., Gyu Kim, B., and Joong-Hoon, A. (2006) Glycosylation of flavonoids with a glycosyltransferase from *Bacillus cereus*. *FEMS Microbiol. Lett.* **258**, 263–268
35. Finel, M., and Kurkela, M. (2008) The UDP-glucuronosyltransferases as oligomeric enzymes. *Curr. Drug Metab.* **9**, 70–76
36. Yuan, L. M., Gao, Z. Z., Sun, H. Y., Qian, S.-N., Xiao, Y. S., Sun, L. L., *et al.* (2016) Inter-isoform hetero-dimerization of human UDP-Glucuronosyltransferases (UGTs) 1A1, 1A9 and 2B7 and impacts on glucuronidation activity. *Sci. Rep.* **6**, 3445
37. Meisburger, S. P., Taylor, A. B., Khan, C. A., Zhang, S., Fitzpatrick, P. F., and Ando, N. (2016) Domain movements upon activation of phenylalanine hydroxylase characterized by crystallography and chromatography-coupled small-angle X-ray scattering. *J. Am. Chem. Soc.* **138**, 6506–6516
38. Schneidman-Duhovny, D., Hammel, M., Tainer, J. A., and Sali, A. (2016) FoXS, FoXSDock and MultiFoXS: single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles. *Nucleic Acids Res.* **44**, W424–W429
39. Holm, L. (2020) Using Dali for protein structure comparison. *Methods Mol. Biol.* **2112**, 29–42
40. Krissinel, E., and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2256–2268
41. 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
42. Bolam, D. N., Roberts, S., Proctor, M. R., Turkenburg, J. P., Dodson, E. J., Martinez-Fleites, C., *et al.* (2007) The crystal structure of two macrolide glycosyltransferases provides a blueprint for host cell antibiotic immunity. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 5336–5341
43. Du, X., Li, Y., Xia, Y. L., Ai, S. M., Liang, J., Sang, P., *et al.* (2016) Insights into protein-ligand Interactions: mechanisms, models, and methods. *Int. J. Mol. Sci.* **17**, 144
44. Boix, E., Swaminathan, G. J., Zhang, Y., Natesh, R., Brew, K., and Acharya, K. R. (2001) Structure of UDP complex of UDP-galactose:β-Galactoside-α-1,3-galactosyltransferase at 1.53-Å resolution reveals a conformational change in the catalytically important C terminus. *J. Biol. Chem.* **276**, 48608–48614
45. Ramakrishnan, B., Balaji, P. V., and Qasba, P. K. (2002) Crystal structure of β1,4-galactosyltransferase complex with UDP-Gal reveals an oligosaccharide acceptor binding site. *J. Mol. Biol.* **318**, 491–502
46. An, J., Kim, S. H., Bahk, S., Vuong, U. T., Nguyen, N. T., Do, H. L., *et al.* (2021) Naringenin induces pathogen resistance against. *Front. Plant Sci.* **12**, 672552
47. Zhang, X., Ran, W., Li, X., Zhang, J., Ye, M., Lin, S., *et al.* (2022) Exogenous application of gallic acid induces the direct defense of tea plant against. *Front. Plant Sci.* **13**, 833489
48. Mierziak, J., Kostyn, K., and Kulma, A. (2014) Flavonoids as important molecules of plant interactions with the environment. *Molecules* **19**, 16240–16265
49. George Thompson, A. M., Iancu, C. V., Neet, K. E., Dean, J. V., and Choe, J. Y. (2017) Differences in salicylic acid glucose conjugations by UGT74F1 and UGT74F2 from *Arabidopsis thaliana*. *Sci. Rep.* **7**, 46629
50. Wetterhorn, K. M., Newmister, S. A., Caniza, R. K., Busman, M., McCormick, S. P., Berthiller, F., *et al.* (2016) Crystal structure of Os79 (Os04g0206600) from *Oryza sativa*: a UDP-glucosyltransferase involved in the detoxification of deoxynivalenol. *Biochemistry* **55**, 6175–6618

51. Graham, T. L. (1998) Flavonoid and flavonol glycoside metabolism in Arabidopsis. *Plant Physiol. Biochem.* **36**, 135–144
52. De Rouck, S., İnak, E., Dermauw, W., and Van Leeuwen, T. (2023) A review of the molecular mechanisms of acaricide resistance in mites and ticks. *Insect Biochem. Mol. Biol.* **159**, 103981
53. Van Leeuwen, T., and Dermauw, W. (2016) The molecular evolution of xenobiotic metabolism and resistance in chelicerate mites. *Annu. Rev. Entomol.* **61**, 475–498
54. Safaa, K., Arriaza, R. H., Kriti, K., Andrea, O. M., Vojislava, G., Miodrag, G., *et al.* (2023) Current status of structural studies of proteins originating from Arachnida. *Syst. Appl. Acarol.* **28**, 298–308
55. Atkins, W. M. (2020) Mechanisms of promiscuity among drug metabolizing enzymes and drug transporters. *FEBS J.* **287**, 1306–1322
56. Radominska-Pandya, A., Bratton, S. M., Redinbo, M. R., and Miley, M. J. (2010) The crystal structure of human UDP-glucuronosyltransferase 2B7 C-terminal end is the first mammalian UGT target to be revealed: the significance for human UGTs from both the 1A and 2B families. *Drug Metab. Rev.* **42**, 133–144
57. Wilkins, M. R., Gasteiger, E., Bairoch, A., Sanchez, J. C., Williams, K. L., Appel, R. D., *et al.* (1999) Protein identification and analysis tools in the ExPASy server. *Methods Mol. Biol.* **112**, 531–552
58. Daneshian, L., Renggli, I., Hanaway, R., Offermann, L. R., Schlachter, C. R., Hernandez Arriaza, R., *et al.* (2022) Structural and functional characterization of β-cyanoalanine synthase from *Tetranychus urticae*. *Insect Biochem. Mol. Biol.* **142**, 103722
59. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **6**, 307–326
60. Vagin, A., and Teplyakov, A. (1997) MOLREP: an automated program for molecular replacement. *J. Appl. Crystallogr.* **30**, 1022–1025
61. Minor, W., Cymborowski, M., Otwinowski, Z., and Chruszcz, M. (2006) HKL-3000: the integration of data reduction and structure solution–from diffraction images to an initial model in minutes. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 859–866
62. Cowtan, K. (2006) The Buccaneer software for automated model building. Tracing protein chains. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 1002–1011
63. Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., *et al.* (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 355–367
64. Painter, J., and Merritt, E. A. (2006) TLSMD web server for the generation of multi-group TLS models. *J. Appl. Crystallogr.* **39**, 109–111
65. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132
66. Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., *et al.* (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, W375–W383
67. Berman, H., Henrick, K., and Nakamura, H. (2003) Announcing the worldwide Protein Data Bank. *Nat. Struct. Biol.* **10**, 980
68. The PyMOL Molecular Graphics System, Version 2.0. (2002). Schrödinger, LLC
69. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., *et al.* (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612
70. Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., and Barton, G. J. (2009) Jalview version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191
71. Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., *et al.* (2016) ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* **44**, W344–W350
72. Consortium, U. (2021) UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* **49**, D480–D489
73. Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797
74. 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
75. Pieper, U., Eswar, N., Webb, B. M., Eramian, D., Kelly, L., Barkan, D. T., *et al.* (2009) MODBASE, a database of annotated comparative protein structure models and associated resources. *Nucleic Acids Res.* **37**, D347–D354
76. Wallace, A. C., Laskowski, R. A., and Thornton, J. M. (1995) LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.* **8**, 127–134
77. Fischer, N., Seo, E. J., and Efferth, T. (2018) Prevention from radiation damage by natural products. *Phytomedicine* **47**, 192–200
78. Hopkins, J. B., Gillian, R. E., and Skou, S. (2017) BioXTAS RAW: improvements to a free open-source program for small-angle X-ray scattering data reduction and analysis. *J. Appl. Crystallogr.* **50**, 1545–1553
79. Case, D. A., Aktulga, H. M., Belfon, I. Y., Ben-Shalom, J. T., Berryman, S. R., Brozell, D. S., *et al.* (2022) *Amber 2022*, University of California, San Francisco
80. Götz, A. W., Williamson, M. J., Xu, D., Poole, D., Le Grand, S., and Walker, R. C. (2012) Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. Generalized born. *J. Chem. Theory Comput.* **8**, 1542–1555
81. Salomon-Ferrer, R., Götz, A. W., Poole, D., Le Grand, S., and Walker, R. C. (2013) Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **9**, 3878–3888
82. Svergun, D. I. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* **25**, 495–503
83. Manalastas-Cantos, K., Konarev, P. V., Hajizadeh, N. R., Kikhney, A. G., Petoukhov, M. V., Molodenskiy, D. S., *et al.* (2021) *ATSAS 3.0*: expanded functionality and new tools for small-angle scattering data analysis. *J. Appl. Crystallogr.* **54**, 343–355
84. Notredame, C., Higgins, D. G., and Heringa, J. (2000) T-coffee: a novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **302**, 205–217
85. Corsi, C., Veronesi, F., Lamberti, C., Bardo, D. M., Jamison, E. B., Lang, R. M., *et al.* (2009) Automated frame-by-frame endocardial border detection from cardiac magnetic resonance images for quantitative assessment of left ventricular function: validation and clinical feasibility. *J. Magn. Reson. Imaging* **29**, 560–568
86. Dmitri, I. S. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* **25**, 495–503
87. Rambo, R. P., and Tainer, J. A. (2013) Accurate assessment of mass, models and resolution by small-angle scattering. *Nature* **496**, 477–481
88. Piiadov, V., Ares de Araújo, E., Oliveira Neto, M., Craievich, A. F., and Polikarpov, I. (2019) SAXSMoW 2.0: online calculator of the molecular weight of proteins in dilute solution from experimental SAXS data measured on a relative scale. *Protein Sci.* **28**, 454–463