Promiscuous yet specific: a methionine-aromatics interaction drives the reaction scope of the family 1 glycosyltransferase GmUGT88E3 from soybean.

Ruben M. de Boer, Dan E. H. Hvid, Elisa Davail, Dovydas Vaitkus, Jens Ø. Duus, Ditte H. Welner, David Teze

AUTHOR ADDRESS. The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark. Department of Chemistry, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark.

Supporting Information Placeholder [MM plots, NMR spectra, HPLC chromatograms, MD snapshots]

ABSTRACT: Family 1 glycosyltransferases (GT1s, UGTs) catalyze the regioselective glycosylation of natural products in a single step. We identified GmUGT88E3 as a particularly promising biocatalyst, able to produce a variety of pure, single glycoside products from polyphenols with high chemical yields. We investigated this particularly desirable duality towards specificity, i.e., promiscuous towards acceptors while regiospecific. Using high-field NMR, kinetic characterization, molecular dynamics simulations and mutagenesis studies, we uncovered that the main molecular determinant of GmUGT88E3 specificity is a methionine-aromatic bridge, an interaction often present in protein structures but never reported for enzyme-substrate interactions.

Glycosides are ubiquitous in nature, yet comparatively absent from our human-managed chemosphere. Indeed, traditional chemistry struggles with glycosylation reactions, requiring multistep synthesis to ensure control of both regio- and stereoselectivity, resulting in high costs and poor atom economy. Conversely, family 1 glycosyltransferases (GT1s, according to the CAZy database classification), are able to catalyze the synthesis of glycosides in a single step from unprotected acceptors.

GT1s are phylogenetically related and present a GT-B fold, with an active side cleft at the interface of the two Rossmann domains. GT1s catalyze inverting reactions using UDP-α-glycosyl as donors (hence they are also commonly termed UDP-dependent glycosyltransferases or UGTs), resulting in β-glycosides. Most of them act via base catalysis with a His-Asp dyad as catalytic residues (Figure 1).

Figure 1. Generalized reaction mechanism for family 1 glycosyltransferases.

Recently, in an effort to develop a predictor for GT1 acceptor selectivity, we assessed reactions catalyzed by 40 GT1s against 32 polyphenol acceptors. Analyzing this dataset, we here identify a high-performing enzyme, GmUGT88E3, which showed great activity (>90% conversion in 24 hours) towards 15/32 acceptors with strict regiospecificity on 7 of those 15. Additionally, when assessing yields above 50% in the same timeframe, it presents strict regiospecificity on 14/32 acceptors. Interestingly, GmUGT88E3 thus appeared to be the most promiscuous enzyme relative to acceptor specificity, while presenting the strictest regioselectivity of all 40 assayed enzymes, two properties that seem opposite. Moreover, while a large number of GT1s present scalability issues related to chemostability, GmUGT88E3 seemed to be unaffected by the process. Overall, GmUGT88E3 appears to be a very promising biocatalyst for the biotechnological chemosynthesis of polyphenols glucosides. Here, we determine the kinetic parameters of these glycosylation reactions and the NMR structures of the products. Then we used molecular dynamics simulations of the Michaelis complexes and mutagenesis studies to identify the molecular determinants of these unusual properties, uncovering that Met127 is forming a methionine-aromatics bridge (Met-Ar) between Phe126 and the polyphenol acceptors. Strikingly, while Met-Ar interactions have been reported as a common feature involved in protein folding and stability, it has never been described in enzyme-substrate interactions.

GmUGT88E3 has been previously investigated. In a study carried out by Livingstone et al. on the transcript levels of six isoflavone 7-O-glucosyltransferases, GmUGT88E3 (GmUGT1/GmIF7GT) transcript levels were found to be significantly higher in early and late seed tissues than the other five GTs. GmUGT88E3 displays activity towards genistein and daidzein, as was shown by Noguchi et al., resulting in the major forms of isoflavones found in soybeans, genistin and daidzin. Though kinetic analysis and structure determination are reported only for genistein, daidzin, and kaempferol, GmUGT88E3 has been described to show activity towards a wide range of acceptors; flavones, flavanones, flavonols, an aurone, a coumarin and a chalcone.
The site of glycosylation highlighted in red was unraveled by NMR spectroscopy (Figs. S3–S8). The reaction mixture that contained product glycoside, UDP, UDP-glucose, and potentially remaining aglycone was lyophilized, dissolved in DMSO-<sup>d6</sup> and directly analyzed by NMR. Glycoside structures of 2,4'-dihydroxybenzophenone and calycosin were determined according to a change in chemical shift of nearby aromatic protons. The sites of glycosylation for the other compounds were identified by 1D NOESY experiments. A response of the nearest aromatic protons was observed through targeted irradiation of the anomeric alpha proton of the resulting glycoside.

**GmUGT88E3** is reported as an isoflavone-7-O-glycosyltransferase, we observed glycosylation at the 7-position for our assayed isoflavones, as well as glycosylation at the 7-position for the flavone Oroxylin A. On the contrary, kaempferol, a flavone, was glycosylated at the 4'-position as was reported by Funaki et al., which raises the hypothesis that in absence of a hydroxide group at the 4'-position, the hydroxide on the 7-position is preferred. Both methyl caffeate and 2,4'-dihydroxybenzophenone are glycosylated at the para position of the aromatic ring. The structure of the 4'-methoxyresveratrol was not determined by NMR since the hydroxides are chemically identical due to free rotation.

In order to establish a better understanding of the determinants governing the broad substrate range of GmUGT88E3, molecular dynamics (MD) simulations of the ternary complexes enzyme:UDP-Glc:acceptor were carried out (Figures 3 and S9–S17). Overall, five amino residues have been identified that form important interactions with the aglycones. First, the catalytic base His15, directed by Asp125, forms a hydrogen bond with the reactive hydroxyl. The only other key residue that appears to interact with all compounds is Met127. While rarely described as a key interactor, methionine is known to act as a bridging motif between aromatic amino acids, appearing in approximately one-third of all known protein structures as a stabilizing motif. In all our simulations (Figures 3 and S9–S16), we observed optimal distances and angles for a Met-Ar bridge between the various polyphenol acceptors and Phe126. Note that these three residues are conssecutive: Asp125, Phe126, and Met127. Besides, Thr155 appears to interact with hydroxides on the opposite side of the glycosylated position in isoflavones. Interestingly, Oroxylin A which is devoid of such a hydroxyl group presents a significantly higher <i>K<sub>m</sub></i> value (Figure 2). Moreover, for 2,4'-dihydroxybenzophenone we observe a strong interaction between Glu197 and the hydroxide on the 4-position, which might explain its low <i>K<sub>m</sub></i> value.
Figure 3. Molecular simulation of the ternary Michaelis complex in the presence of UDP-Glc and 2,4'-dihydroxybenzophenone. Left, snapshots showing key interactions showing 2,4'-dihydroxybenzophenone (cyan), UDP-glucose (yellow), and interacting residues (white). Right, monitoring of selected distances and angles from MD trajectories. Clearly, Met127 appeared to be a key residue within the active site of GmUGT88E3 and is absent from the other 39 GT1s assayed. Therefore, five mutants were constructed to verify the significance of Met127 in substrate binding and regioselectivity. However, M127A, M127L, M127I, and M127E could not be purified, possibly due to the disruption of the methionine-aromatic bridging motif between Phe126 and Met127. Only M127K, which was chosen as it might present a replacement stabilizing cation-π interaction, was successfully purified and found active (Figure S10). However, the activity was drastically reduced compared to the wild-type enzyme, leading to severely reduced yields even at 10-fold higher enzyme loadings (Fig. 4). This large difference in activity clearly demonstrates how the loss of the Met127 is detrimental to activity.

To summarize, the potent GmUGT88E3 was identified as a particularly interesting biocatalyst, being able to deliver high product yields (>90%) and perfect regiospecificity for a broad substrate range of polyphenol glucosylation. A substrate panel was established based on observed regiospecificity, kinetic values and chemical structures were determined. A fast NMR methodology to determine the structure of the products directly from the enzymatic mixture was established via 1D NOESY experiments. Strikingly, the molecular determinant that governs selectivity appears to be a Met-Ar bridge, leading to a Phe126-Met127-phenol interaction.

Figure 4. Comparison of GmUGT88E3 wild-type (2 µg/mL) and mutant GmUGT88E3_M127K (20 µg/mL) against 6 of the assessed aglycones. The reactions were monitored by HPLC after 90 min at 25°C in Na-phosphate buffer pH 8. Similar results after 30 min are displayed in Fig. S18.

ASSOCIATED CONTENT
Supporting Information
Michaelis-Menten plots, representative HPLC chromatograms, supporting figures for glycoside structure as determined by NMR, supporting figures for active site analysis with MD data, and additional experimental details, materials and methods (PDF).

AUTHOR INFORMATION
Corresponding Author
David Teze
dt@chem.ku.dk, david.teze@gmail.com
Ditte Hededam Welner
diwel@biosustain.dtu.dk

Present Addresses
† Department of Chemistry, University of Copenhagen. Universitetsparken 5, DK-2100 Copenhagen O, Denmark.
The authors declare no competing financial interests.

ACKNOWLEDGMENTS

The authors thank Kasper Enemark-Rasmussen for technical assistance. This work was supported by the Novo Nordisk Foundation (grants NNF18OC0034744, NNF10CC1016517, and NNF20CC0035580 to DHW), and the Villum Foundation (DTU NMR Center, to JD).

REFERENCES


(10) Farmakalidis, E.; Murphy, P. A. Oestrogenic Response of the CD-1 Mouse to the Soy-Bean Isoflavones Genistein, Genistin and Daidzir; 1984; Vol. 22.
