Predicting the substrate specificity of a glycosyltransferase implicated in the production of phenolic volatiles in tomato fruit

Thomas Louveau[1,5,*], Celine Leitao[1,6,*], Sol Green[2,*], Cyril Hamiaux[2], Benoît van der Rest[1], Odile Dechy-Cabaret[3,4], Ross G. Atkinson[2] and Christian Chervin[1]

1 Université de Toulouse, UMR Génomique et Biotechnologie des Fruits, INRA-INP/ENSAT, Castanet-Tolosan, France
2 The New Zealand Institute for Plant & Food Research Ltd, Auckland, New Zealand
3 CNRS, LCC (Laboratoire de Chimie de Coordination), Toulouse, France
4 Université de Toulouse, UPS, INP, LCC, Toulouse, France
5 John Innes Centre, Dep. Metabolic Biology, Norwich, UK
6 Université de Strasbourg, Equipe de Chimie Analytique des Molécules Bioactives, Faculté de Pharmacie, Illkirch, France

Keywords
aroma; docking; eugenol; guaiacol; isosalicin; methyl salicylate

Correspondence
C. Chervin, ENSAT, BP 32607, 31326 Castanet-Tolosan, France
Fax: +33 5 3432 3873
Tel: +33 5 3432 3870
E-mail: chervin@ensat.fr

*These authors contributed equally to this work

The volatile compounds that constitute the fruit aroma of ripe tomato (*Solanum lycopersicum*) are often sequestered in glycosylated form. A homology-based screen was used to identify the gene *SlUGT5*, which is a member of UDP-glycosyltransferase 72 family and shows specificity towards a range of substrates, including flavonoid, flavanols, hydroquinone, xenobiotics and chlorinated pollutants. *SlUGT5* was shown to be expressed primarily in ripening fruit and flowers, and mapped to chromosome I in a region containing a QTL that affected the content of guaiacol and eugenol in tomato crosses. Recombinant SlUGT5 protein demonstrated significant activity towards guaiacol and eugenol, as well as benzyl alcohol and methyl salicylate; however, the highest *in vitro* activity and affinity was shown for hydroquinone and salicyl alcohol. NMR analysis identified isosalicin as the only product of salicyl alcohol glycosylation. Protein modelling and substrate docking analysis were used to assess the basis for the substrate specificity of SlUGT5. The analysis correctly predicted the interactions with SlUGT5 substrates, and also indicated that increased hydrogen bonding, due to the presence of a second hydrophilic group in methyl salicylate, guaiacol and hydroquinone, appeared to more favourably anchor these acceptors within the glycosylation site, leading to increased stability, higher activities and higher substrate affinities.

Introduction

Tomato (*Solanum lycopersicum*) aroma is a key factor that determines fruit quality and consumer acceptability. The volatile compounds contributing to tomato aroma increase during fruit ripening, peaking at mature breaker or mature red stages. Over 400 volatile compounds have been identified in tomato fruit [1], with recent studies showing that there is a significant variation between cultivars [2,3]. These and previous studies [4,5] showed that most aroma compounds are stored as glycosides. The proportion of glycosides found in various cultivars is also very variable, with proportions of glycosides of benzyl alcohol, eugenol,

Abbreviations
GT, glycosyltransferase; PSPG, plant secondary product glycosyltransferase; SlUGT5, *Solanum lycopersicum* UDP-glycosyltransferase 5; UGT, UDP-GlycosylTransferase.

guaiacol and methyl salicylate varying from 49–88%, 36–68%, 6–50% and 42–73%, respectively, of the corresponding aglycone [2,3]. Glycosides contributing to tomato aroma also tend to accumulate in fruit over the ripening phase [2].

The glycosylation of aroma volatiles is usually catalysed by glycosyltransferases (GTs), which mediate the transfer of a sugar residue from an activated nucleotide sugar to acceptor molecules. Many GTs have been characterized in the plant kingdom, and this family of enzymes has been the subject of several reviews [6,7]. All plant GTs contain a common signature motif of 44 amino acids, known as the plant secondary product glycosyltransferase box (PSPG) [7], which is thought to be involved in binding the UDP moiety of the activated sugar. Phylogenetic analysis [8] has classified plant UDPglycosyltransferase (UGT)1 sequences into 29 families (UGT71–UGT99) comprising 14 groups (A–N). This classification allows rapid integration of newly cloned GTs into existing trees. In tomatoes, GT activity in extracts partially purified using ammonium sulfate has been shown to increase over the ripening phase [9]. Although there are no reports showing the direct involvement of UGTs in the glycosylation of tomato aroma volatile precursors, several GTs from other plant species have been shown to accept known tomato aroma compounds as substrates. For example, eugenol is glycosylated by an arbutin synthase of *Rauvolfia serpentina* [10], UDP-glucose:*p*-hydroxymandelonitrile-O-glucosyltransferase from *Sorghum bicolor* catalyses the glycosylation of geraniol and benzyl alcohol [11], and AtSAGT1 from *Arabidopsis thaliana* can catalyze the *in vitro* formation of methyl salicylate glucose from methyl salicylate [12].

UGTs were initially thought to be promiscuous enzymes; however, the substrate specificity of UGTs appears to be limited by regio-selectivity [13,14], and in some cases UGTs have been shown to be highly specific [15,16]. Our understanding of the glycosylation mechanism and how substrate preference is determined has been greatly improved by the publication of crystal structures for five plant UGTs [17–19]. Despite relatively low levels of sequence conservation, all plant UGTs have very similar structures, in which the two domains (N- and C-terminal, both adopting Rossman-like folds) form a cleft to accommodate the substrates, nucleotide sugar and acceptor. Family 1 GTs are inverting enzymes that invert the anomeric configuration of their catalytic products compared to their respective substrates [17,18]. Family 1 GT-mediated glycosylation occurs through a direct-displacement, S_N2-like, mechanism, whereby a highly conserved catalytic histidine acts as a general base to abstract a proton from the acceptor substrate, allowing nucleophilic attack on the C1 atom of the UDP-sugar to form the glycosylated product [17–19]. Despite this information, it is very difficult to predict GT substrate preference based on structural characteristics alone.

In this study, we characterize a tomato GT that shows activity towards aglycones associated with tomato fruit aroma, and use substrate docking analysis to assess the basis for the substrate specificity.

Results and Discussion

Cloning and sequence analysis of *SlUGT5*

The SGN Unigene Database (http://solgenomics.net/) was searched for tomato UGT sequences with similarity to FaGT2, a UDP-glucose-cinnamate glucosyltransferase involved in the accumulation of cinnamoyl-D-glucose during fruit ripening in strawberry (*Fragaria* × *ananassa*), a precursor of volatiles linked to strawberry aroma (accession number Q66PF4) [20]. A total of 121 putative UGT unigenes were initially identified, of which 34 had expression profiles described in the Tomato Functional Genomics Database (http://ted.bti.cornell.edu). Four of these 34 unigenes (SGN-U315028, SGN-U312947, SGN-U316027 and SGN-U313478) were highly expressed during fruit ripening, either in wild-type fruit or in the *never-ripe* mutant (data not shown). In a preliminary study, these four genes were cloned, fully sequenced (Fig. S1) and expressed in *Escherichia coli* with an N-terminal polyhistidine tag. The protein corresponding to the SGN-U315028 unigene was soluble (Fig. S2) and active, and was therefore chosen for further detailed phylogenetic and biochemical analysis.

The full-length ORF corresponding to SGN-U315028 (named SlUGT5) was 1476 bp long, and encoded a protein with a predicted molecular mass of 54.1 kDa and a pI of 5.63. The sequence contained the PSPG consensus sequence of 44 amino acids found in all plant UGTs (Fig. S3). A phylogenetic comparison using SlUGT5 and members of the published Arabidopsis UGT tree [8,21] indicated that the tomato sequence clustered most closely with UGT72B family members in group E (Fig. 1). On this basis, SGN-U315028 was designated SlUGT72B (*Solanum lycopersicum* UDP-glycosyltransferase 72B).

SlUGT5 displayed highest amino acid identity (83%) to an uncharacterized protein from *Lycium barbarum* (BAG80556) and HpUGT72B11 from *Hieracium pilosella* (ACB56923), a glucosyltransferase that acts on flavonoids and flavonols [22]. In the UGT72B family, two other UGTs have defined substrate preferences – an

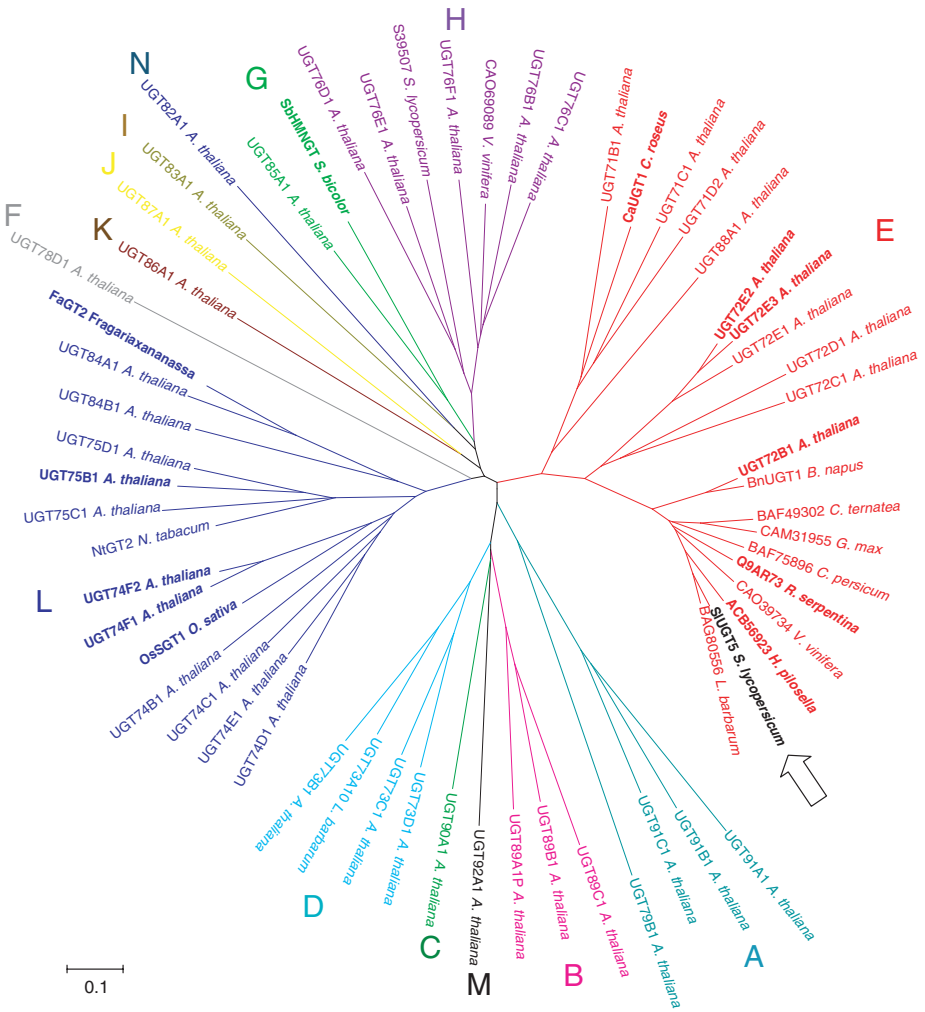

Fig. 1. Phylogenetic relationship of SlUGT5 from *Solanum lycopersicum* (HM209439) with other members of plant glycosyltransferase family 1 (according to the Carbohydrate-Active enZymes, CAZy, data base). Groups A–N have been defined previously [8,21]. The unrooted tree was constructed using MEGA 4 after alignment of sequences using Clustal W2. Arabidopsis UGT amino acid sequences were obtained from http://www.p450.kvl.dk/UGT.shtml. The other genes are: BAG80556 from *Lycium barbarum* (B6EWZ3); ACB56923 glucosyltransferase HpUGT72B11 from *Hieracium pilosella* (B2CZL2); CAO39734 and CAO69089 from *Vitis vinifera*; BAF75896 from *Cyclamen persicum*; Q9AR73 arbutin synthase from *Rauvolfia serpentina*; CAM31955 from *Glycine max* (A5I866); BAF49302 from *Clitoria ternatea* (A4F1R9); 3,4-dichlorophenol glycosyltransferase BnUGT2 from *Brassica napus* (A5I865); salicylic acid glucosyltransferase OsSGT1 from *Oryza sativa* (Q9SE32); cinnamate glycosyltransferase FaGT2 from *Fragaria* × *ananassa* (Q66PF4); *p*-hydroxymandelonitrile glucosyltransferase SbHMNGT from *Sorghum bicolor* (Q9SBL1); UGT73A10 from *Lycium barbarum* (B6EWX3); NtGT2 from *Nicotiana tabacum* (Q8RU71); S39507 glucuronosyl transferase from *Solanum lycopersicum* (S39507); CaUGT1 from *Catharanthus roseus* (Q6F4D6). Accesion numbers for SwissProt (UniProtKB/TrEMBL) are given in brackets.

arbutin synthase from *R. serpentina* (Q9AR73), which shows maximal activity toward hydroquinone and acts on xenobiotics [10], and a bifunctional O- and N-glucosyltransferase from *Arabidopsis thaliana* UGT72B1) that can detoxify the chlorinated pollutants trichlorophenol and dichloroaniline [23–26]. In the closely related UGT72E family, three genes from *A. thaliana* (UGT72E1, 2 and 3) have been shown to play an important role in the synthesis of monolignols [27,28]. UGT72L1 may be involved in the production of epicatechin 3′-*O*-glucoside in the *Medicago truncatula* seed coat [29]. An alignment of SlUGT5 with related group E UGT sequences is shown in Fig. S3.

Mapping and expression analysis of *SlUGT5*

Using the recently assembled tomato genomic sequence (http://solgenomics.net/), *SlUGT5* was shown to be located 41 kbp upstream of the TG650 marker, which maps to chromosome I (located at 88.5 cM according

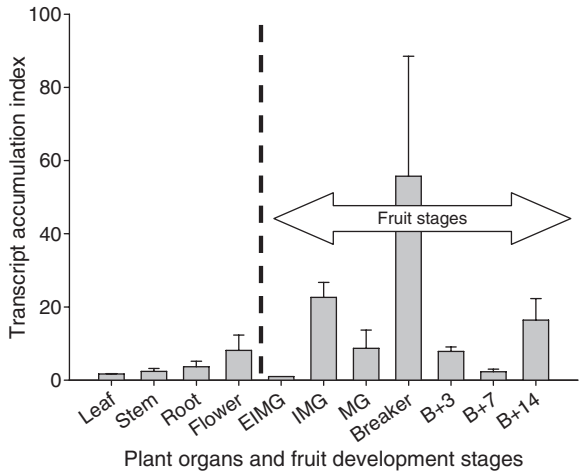

Fig. 2. *SlUGT5* mRNA accumulation profile in tomato plant organs. Fruit development stages: EIMG, IMG and B+ '×' indicate early immature green, immature green and breaker plus '×' days, respectively. The transcript accumulation index was calculated using actin as a reference gene, and the EIMG value was set at 1. Error bars represent the standard error with $n = 3$ biological replicates.

to the Tomato-EXPEN 2000 map). Interestingly, this region of chromosome I has been shown to contain a QTL affecting the content of guaiacol and eugenol in crosses between cherry tomatoes and three independent large-fruit cultivars [30]. The importance of this region was confirmed in flavour-related metabolite profiling in *Solanum penellii* derived introgression lines (IL) (http://ted.bti.cornell.edu). The IL 1-2 line carrying the *S. pennelli* chromosome I segment containing *SlUGT5* has dramatically reduced methyl salicylate and methyl benzoate content compared to other IL lines.

The mRNA accumulation profile of *SlUGT5* in a range of tomato vegetative and fruit tissues was examined by quantitative PCR (Fig. 2). Low transcript levels of SlUGT5 were measured in stem, leaves and roots, but there was some transcript accumulation in flowers Transcripts accumulated to higher levels in fruit from the immature green stage to 14 days after breaker stage (fully ripe). There was some variability in *SlUGT5* transcript accumulation in developing and senescing fruit, with immature green, breaker and breaker + 14 day stages accumulating more transcript. The observed trend, of an increase up to the breaker stage and then a decrease, matches the results observed in microarray data available from the Tomato Functional Genomics Database (Table S1). Although there were no obvious physical differences in the plants and fruit examined, we cannot exclude the possibility that the late transcript increase at breaker + 14 days could be due to fungal infection.

Table 1. V_{max} (nkat·mg^{-1} protein), relative velocities (V_{rel}) and K_m (mM) of SlUGT5 at pH 7.5 in the presence of UDP-glucose (10 mM) for acceptors known to be involved in tomato aroma.

Substrate	V_{max}	V_{rel}	K_m
Methyl salicylate	22.1	100	2.3
Guaiacol	19.8	90	10.2
Eugenol	7.62	34	1.1
Benzyl alcohol	4.43	20	62.3
Phenyl ethanol	Not detected	–	–

Indeed, it has been observed previously (Table S1) that *SlUGT5* expression is induced 36 or 60 h after plant infection with the pathogen oomycete *Phytophthora infestans*, and that this induction coincides with the expression of pathogen-related proteins and salicylic acid synthesis during hypersensitive response initiation [31].

Recombinant enzyme activity

The mapping and expression data suggested that SlUGT5 might have a role in glycosylating aroma compounds during tomato fruit ripening. To determine the substrate specificity of SlUGT5, recombinant protein was expressed in *E. coli* and purified using a cobalt affinity resin. The activity of the recombinant protein was firstly tested against a range of hydroxyl benzyl alcohols commonly found as glycosides in tomatoes [2,3,5]. In the presence of UDP-glucose, SlUGT5 showed activity with methyl salicylate, guaiacol, eugenol and benzyl alcohol (Table 1), but no activity was detected with phenyl ethanol or salicylic acid. The products of the glycosylation reaction were analysed by LC-MS for methyl salicylate, guaiacol, eugenol and benzyl alcohol (Fig. S4). ESI-MS analysis in positive mode (presence of sodium adduct at $m/z = M + 23$) showed that the major product in all cases was the corresponding monoglycoside.

Similar substrates have previously been shown to be used by other UGTs in family 72 (e.g. the arbutin synthase of *R. serpentina* (Q9AR73) uses eugenol and methoxyphenols, which are close in structure to guaiacol). The activity of SlUGT5 was then tested with other compounds that have been shown to be substrates of HpUGT72B11 of *H. pilosella* (ACB56923) and the arbutin synthase of *R. serpentina*. SlUGT5 had a K_m for both hydroquinone and salicyl alcohol comparable to that for eugenol and methyl salicylate (Table 2). SlUGT5 also accepted kaempferol and cinnamyl alcohol as substrates, with 10 and 2% of the activity of hydroquinone, respectively (data not shown). The relative activities for hydroquinone and kaempferol differ

Table 2. V_{max} (nkat·mg^{-1} protein), relative velocities (V_{rel}) and K_m (mM) of SlUGT5 for acceptors used by related UGT enzymes.

Substrate	V_{max}	V_{rel}	K_m
Hydroquinone	121.3	100	0.54
Salicyl alcohol	77.5	64	0.9
4-OH benzyl alcohol	47.3	39	10

considerably from those of HpUGT72B11 reported for the same substrates in a previous study [22]. SlUGT5 activity showed a temperature optimum of 37–40 °C and a pH optimum of 7.5 for both benzyl alcohol and salicyl alcohol.

The glycoside produced by the SlUGT5 using salicyl alcohol showed a different retention time (approximately 10 min, Fig. S4) to that of a β-salicin standard run under the same conditions (v 9 min, data not shown). More detailed analysis using NMR was performed to identify the product of the reaction. The regio-selectivity of the enzymatic glucosylation using salicyl alcohol was analysed using preparative liquid chromatography and NMR. ^{1}H and ^{13}C-NMR analyses were performed in D_2O, and compared to NMR data for the four salicin isomers β-salicin [32], β-isosalicin [33], α-salicin [34,35] and α-isosalicin [34], previously reported in the literature (see Fig. S5). The ^{1}H-NMR spectrum included a doublet signal at 4.47 ppm attributable to a β-anomeric proton of the glucosyl moiety, as this signal had a large coupling constant (J = 8.1 Hz). Moreover, the carbon signal of C7 (67.0 ppm) was de-shielded compared to salicyl alcohol (60.1 ppm) [34] or natural β-salicin (59.2 ppm) under the same conditions (D_2O), indicating that the glucose moiety is attached to the hydroxyl group at C7 rather than C1. These results identify the purified product as β-isosalicin, indicating that the glycosylation of salicyl alcohol catalysed by the purified enzyme proceeds in a both regio-selective (isosalicin and not salicin) and stereo-selective (only the β-anomer) manner. In the study of arbutin synthase (Q9AR73) of *R. serpentina*, the authors showed that saligenin (salicyl alcohol) was accepted as a substrate, but the selectivity was not checked [10].

UDP-galactose and UDP-glucuronate were tested as alternative activated sugar donors, with salicyl alcohol as an acceptor. The K_m for UDP-galactose was similar to that for UDP-glucose (0.31 versus 0.9 mM, respectively), but its V_{max} was lower than that observed for UDP-glucose (0.44 versus 77.5 nkat·mg^{-1}, respectively). No activity was detected when UDP-glucuronate was used as the donor. SlUGT5 can therefore be designated as a UDP-glycosyltransferase, utilizing UDP-glucose and UDP-galactose as its preferred activated sugar donors.

Protein modelling

To understand the basis for the substrate specificity of SlUGT5 (Tables 1 and 2), a SlUGT5 protein homology model was constructed using Modeller 9.7 [36], with the crystal structure of Arabidopsis UGT72B1 (60.5% identity) as the template. In the crystal structure of the UGT72B1 Michaelis complex with the oxygen acceptor 2,4,5-trichlorophenol and a non-transferable UDP-glucose analogue (UDP-2-deoxy-fluoroglucose), the acceptor lies in the binding pocket with its hydroxyl group hydrogen-bonded to the catalytic histidine, in perfect position for nucleophilic attack on the C1 atom of the glucose [26]. No additional interaction between the acceptor and the surrounding proteins atoms of the binding pocket was observed [26]. Compared to other plant UGTs, members of family 72 are characterized by an additional loop in the C-terminal domain comprising 16 or 17 residues (Ser306–Pro324 in UGT72B1) (Fig. S3). In the Arabidopsis UGT72B1 structure, an interaction between Tyr315 and the main-chain atoms of Ser14 and Pro15 anchors this loop within the vicinity of the active site, therefore significantly reducing the size and accessibility of the acceptor binding pocket (Fig. S6). In SlUGT5, this tyrosine is replaced by a phenylalanine (Phe311), suggesting that local rearrangement of the long additional loop covering the opening of the binding pocket may occur.

Docking experiments were initially performed using methyl salicylate, guaiacol, eugenol, benzyl alcohol and phenyl ethanol. For each of these compounds, 50 independent acceptor binding conformations (solutions) were generated, and a range of potential binding clusters was obtained. In each case, at least two clusters were consistent with the geometry required to support nucleophilic attack on the glucose C1 atom (Fig. 3A–E). Interestingly, the alternative binding clusters obtained for eugenol showed an increase in non-productive catalytic outcomes (34/50) compared to those observed when methyl salicylate (13/50) or guaiacol (1/50) were docked into the SlUGT5 active site. These findings are consistent with the decreased SlUGT5 activity (V_{max}) in the presence of eugenol (Tables 1 and 2). The predicted binding conformations for benzyl alcohol and phenylethanol all have the alcohol hydroxyl positioned in a manner consistent with UGT activity, but SlUGT5 shows low activity and binding affinity for benzyl alcohol and no detectable activity towards phenylethanol. Compared to methyl salicylate, guaiacol and eugenol, the most notable difference in the docking of phenylethanol (Fig. 3D) and benzyl alcohol (Fig. 3E) was that their interactions with the catalytic histidine and glucose C1 atom could

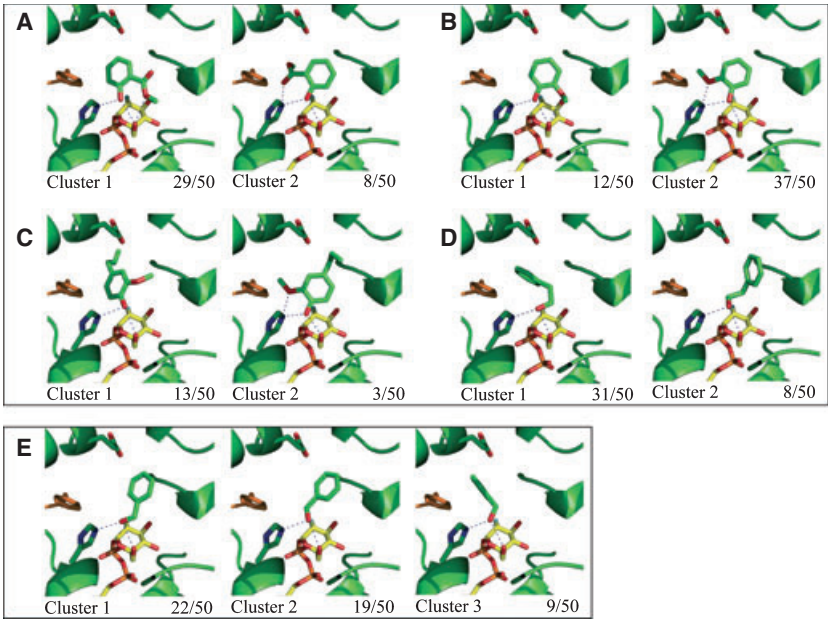

Fig. 3. Docking of methyl salicylate (A), guaiacol (B), eugenol (C) phenylethanol (D) and benzyl alcohol (E) in the SlUGT5 model. One molecule representative of each binding cluster is shown in all cases. The number of acceptor binding conformations (solutions) associated with each cluster is expressed as a fraction of the 50 solutions generated from the docking analysis. Acceptor binding conformations that are not catalytically relevant are not shown. The catalytic residues His17, Glu81 and Phe311 are represented in stick mode, with Phe311 shown in orange. Hydrogen bonds between the docked acceptor molecules and protein atoms are represented as dashed lines. The approximate free binding energies and kI values for all binding clusters are given in Table S2.

only sustain a maximum of two hydrogen bonds, compared to three hydrogen-bond interactions with methyl salicylate and guaiacol (Fig. 3A,B respectively). The decreased hydrogen bonding capacity of benzyl alcohol and phenylethanol could affect their ability to maintain catalytically favourable binding geometries.

Docking of hydroquinone in the acceptor binding pocket of SlUGT5 resulted in a single conformation cluster (Fig. 4A) in which the alcohol hydroxyl group was suitably positioned for nucleophilic attack. This positioning was further strengthened via the second hydroxyl group, which interacts with Glu81 at the other end of the binding pocket (Fig. 4A). As Glu81 (Glu83 in UGT72B1) is strictly conserved within family 72 UGTs (Fig. S3), this conformation provides a structural basis for the high activity of SlUGT5 (Tables 1 and 2) and arbutin synthase [10] for hydroquinone. On the assumption that interaction between Glu81 and a second acceptor hydroxyl group translates to increased UGT activity, we predicted that 4-OH benzyl alcohol would bind in a similar manner to hydroquinone (Fig. 4B) and would show higher activity compared to benzyl alcohol as a substrate for SlUGT5. Our results confirmed this prediction, with SlUGT5 showing a sixfold increase in binding affinity for 4-OH benzyl alcohol (K_m of 10 mM) compared with benzyl alcohol (K_m of 62.3 mM) and a

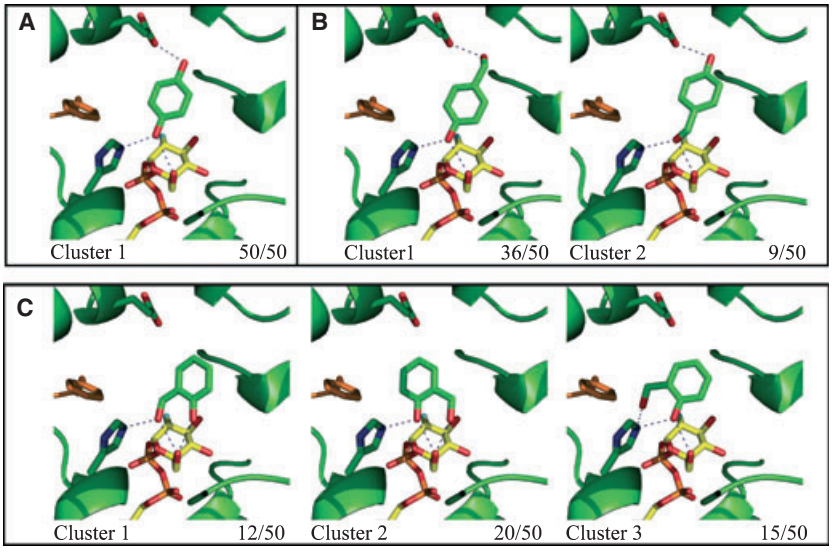

Fig. 4. Docking of hydroquinone (A), 4-OH benzyl alcohol (B) and salicyl alcohol (C) in the SlUGT5 model. Representations of catalytic residues and hydrogen bonds are as for Fig. 3. The free binding energies and kI values for each binding cluster are given in Table S3.

higher activity (V_{max} of 47 nkat·mg^{-1}) compared with benzyl alcohol (V_{max} of 4.4 nkat·mg^{-1}) (Table 2).

SlUGT5 also showed high activity towards salicyl alcohol (Table 2), and NMR analysis identified β-isosalicin as the reaction product. Docking of salicyl alcohol into the acceptor binding pocket yielded three main binding clusters (Fig. 4C). In cluster 1, the primary alcohol hydroxyl group of salicyl alcohol was hydrogen-bonded to the catalytic histidine, and nucleophilic attack on the glucose C1 atom would trigger the formation of β-isosalicin. This conformation is stabilized by an additional hydrogen bond between the phenolic hydroxyl group of salicyl alcohol and the glucose O6 atom. In cluster 2, the situation is reversed, with the phenolic hydroxyl group of salicyl alcohol positioned for attack on the glucose C1, while the primary alcohol hydroxyl group stabilizes the conformation by interacting with the glucose O6 atom. Such a conformation would lead to production of β-salicin rather than β-isosalicin. The third cluster, which shows both the salicyl alcohol hydroxyl groups hydrogen-bonded to the catalytic histidine, could potentially result in either of the salicin isomers being formed. The calculated binding affinities (K_i) for the three clusters are similar (Table S3), and, as such, cannot explain the observed preference for the β-isosalicin production determined by NMR. The main difference between the conformation clusters lies in the position of the aromatic ring of salicyl alcohol in the binding pocket. In clusters 1 and 3, the ring is oriented 'inside', towards the conserved core of the binding pocket, but in cluster 2, it is oriented towards the long loop covering the opening of the binding pocket (Figs 4C and S6), in which most structural variations among UGTs are found. As Tyr315 of Arabidopsis UGT72B1 is replaced by Phe311 in SlUGT5, a structural rearrangement of the long additional loop is likely to occur in SlUGT5 compared to the model. Such rearrangement may modify the shape of the binding pocket to prevent binding of salicyl alcohol in conformation 2, and favour production of the β-isosalicin isomer over β-salicin (Fig. 4C). It is more difficult to determine why cluster 3 would favour β-isosalicin formation, but the exact positioning of the catalytic histidine is likely to be crucial to product outcome.

Conclusions

To our knowledge, this is the first report describing the cloning and characterization of a glycosyltransferase involved in sequestration of tomato aroma compounds as glycosides. SlUGT5 was able to glycosylate methyl salicylate, guaiacol and eugenol, which have all been reported to be present as free volatiles and as glycosides in several tomato cultivars [2,3] and that contribute to consumer perceptions of tomato aroma [1,2]. The expression of *SlUGT5* mRNA during fruit development and ripening is consistent with the SlUGT5 enzyme having a role in the accumulation of glycosides of these compounds during this period. The three other UGT unigenes that we identified may be important in the glycosylation of other key aroma volatiles (e.g. phenyl ethanol) or act to form di- and tri-glycosides [37] during tomato fruit ripening.

Protein homology modelling and substrate docking analysis provided clues to the structural basis for differences in SlUGT5 activity towards the endogenous tomato precursors (methyl salicylate, guaiacol and eugenol) and other substrates tested (hydroquinone and salicyl alcohol). Acceptor substrates possessing two hydrophilic groups generally showed increased activity compared with those with a single hydrophilic substituent. The presence of a second hydrophilic substituent provided an additional hydrogen-bond interaction, and hence was assumed to confer a more stabilized binding configuration. The positioning of the two hydrophilic groups was also important for activity, with *para*-substituted benzene rings being favoured over those that were *ortho*-substituted. There was also good evidence to support the importance of an active-site glutamate residue (Glu81 in SlUGT5; conserved in family 72 UGTs) in determining these preferences by conferring optimal geometry for the single displacement mechanism underlying SlUGT5-mediated glycosylation. The structural insights gained in this study provide a rational basis to test the repertoire of SlUGT5 substrates, and potentially to increase the range of family 72 UGT substrates using a mutagenesis-based approach.

Experimental procedures

Plant material

Tomato *Solanum lycopersicum* (cv. MicroTom) plants were grown in a controlled environment as previously described [38]. Whole fruit were picked at various developmental stages [39] and kept at −80 °C until required. For nucleic acid extraction, batches of five fruit, each from a different plant, were ground under liquid nitrogen using a steel bead grinder (Dangoumau, France).

SlUGT5 cloning and protein purification

The open reading frame (ORF) of *SlUGT5* was amplified from cDNA of immature green, mature green and breaker + 7 days tomato fruits using Gateway® sense primer

G-GT5-F (5′-AAAAAGCAGGCTTCATGGCGCAAATT CCTCATAT-3′) and antisense primer G-GT5-R (5′-AGA-AAGCTGGGTGTCGTGGGCACGATAACGAG-3′). The ORF was then sub-cloned into entry vector pDONR207 (Invitrogen, Karlsruhe, Germany) by introducing the required attB1 and attB2 recombination sites in a two-step PCR process, and recombined into expression vector pDEST™ 17 (Invitrogen) containing a N-terminal polyhistidine tag. The clone was transformed into competent *E. coli* cells (strain BL21-AI; Invitrogen). *E. coli* cells were grown at 37 °C in 100 mL LB medium containing 50 $\mu g \cdot mL^{-1}$ carbenicillin, and expression was induced by 0.2% arabinose for 5 h at 24 °C. The cells were pelleted by centrifugation at 12 000 ***g*** for 10 min, and resuspended in 4 mL of extraction buffer consisting of 20 mM Tris/HCl (pH 8), 500 mM NaCl, 10% v/v glycerol, 0.05% v/v Tween-20, 100 U DNase per mL and 1 mM mercaptoethanol. The cells were disrupted using a bead grinder under liquid nitrogen, then by three cycles of thawing/freezing. The homogenate was incubated at 4 °C for 1 h after addition of a protease inhibitor mix (Roche, Meylan, France), and then centrifuged at 48 000 ***g*** for 20 min at 4 °C. The supernatant was subjected to TALON™ affinity chromatography: 1 mL of supernatant was mixed with 0.3 mL of TALON resin (Clontech/BD Biosciences, Saint-Germain-en-Layr, France) pre-equilibrated three times with extraction buffer without DNase. The recombinant protein was allowed to bind to the resin for 30 min at 4 °C, and, after transfer to a column (a 1 mL pipette tip plugged with glass cotton), the resin was washed twice with 1 mL of extraction buffer, and recombinant protein was specifically eluted with increasing concentrations of imidazole. Protein quantification was performed by Bradford assay (Bio-Rad, Hercules, CA, USA), using bovine serum albumin (BSA) as the standard. Cell lysates and purified protein preparations were separated by SDS/PAGE, and protein bands were visualized using silver staining.

Genetic studies

The NCBI protein BLAST program (http://blast.ncbi.nlm.nih.gov/Blast.cgi) was used to find homologues of SlUGT5 in the Sol Genomics Network (SGN) Unigene database (http://solgenomics.net/). Sequences were aligned using MAFFT (http://www.imtech.res.in/raghava/mafft/). The unrooted phylogenetic tree was constructed using MEGA 4 (http://www.megasoftware.net/) by the neighbor-joining method. Defining the location of the SlUGT5 on chromosome I was performed using Tomato-EXPEN 2000 version 52 (http://solgenomics.net/cview).

Quantitative PCR

RNA extractions were performed using cetyl trimethylammonium bromide (CTAB) [39]. Quantitative PCR was performed as described previously [40] using an optimal primer concentration of 300 nM. All quantitative PCR experiments were run in triplicate using cDNAs synthesized from three biological replicates. Each sample was run in three technical replicates on a 384-well plate. Relative fold differences (transcript accumulation index) were calculated based on the comparative C_t method, using actin as an internal standard, and the $2^{-\Delta\Delta C_t}$, with the highest ΔCt as the basal reference for each gene.

Activity assays and HPLC

SlUGT5 activity assays were performed in 50 mM Tris (pH 7.5), 1 mM $MgCl_2$ at 37 °C. The saturating conditions of donor were determined at 10 mM UDP glucose for 700 ng of SlUGT5 protein in a final volume of 70 µL. Reactions were stopped after 5, 10 and 15 min (linear conditions) by addition of 1/20 v/v trichloroacetic acid at 240 $mg \cdot mL^{-1}$, and immediately transferred to ice. Impurities were eliminated by centrifugation at 13 000 ***g*** (4 min, 4 °C) prior to HPLC analysis.

The analysis of samples corresponding to the enzymatic kinetic reactions was performed by reverse-phase HPLC (HPLC Dionex UltiMate 3000 driven by Chromeleon version 6.80, Voisins-le-Bretonneux, France) on a C18-2 column (Interchim, Montluçon, France, Interchrom Upti-prep Strategy, 100 Å, 5 µm, 150 × 2 mm). The eluents used were H_2O + 0.1% formic acid (eluent A, polar) and acetonitrile (eluent B, non-polar). The mobile phase was constant (2% eluent B) for 2 min at a flow rate of 0.2 $mL \cdot min^{-1}$, then modified linearly as follow: 2–15% eluent B over 3 min, 15–40% eluent B over 7 min, 40–70% eluent B over 1 min, constant flow 70% eluent B over 5 min, linear gradient 70–2% eluent B over 1 min. The injection volume was 10 µL. The detection wavelengths for the substrates and their corresponding glycosides were 303 nm for methyl salicylate, 276 nm for guaiacol and eugenol, 221 nm for benzyl alcohol, 272 nm for salicyl alcohol and 288 nm for hydroquinone. Given that all reactions studied here are equimolar, and that in each case we observed an increase in the product peak only, the activities for each aglycone were calculated from sample substrate and product peak areas, relative to external standards. When running experiments for determination of K_m and V_{max} (calculated from Lineweaver–Burk plots), the reactions were initiated by addition of the aglycone to the reaction tube (t = 0). Control reactions were performed as above using boiled enzymes. The enzyme activities were expressed as nkat of the related glycoside per mg protein, and the K_m was expressed in mM of the relevant substrate.

LC-MS and NMR

LC-MS and NMR analyses were performed to confirm the identity of the products from SlUGT5 *in vitro* activity tests. LC-MS analyses were performed using an Agilent 1100 series

(Massy, France) HPLC under the same LC conditions (column and elution gradient) as in the HPLC analysis. ESI-MS analyses were performed using a Q-Trap mass spectrometer (Applied Biosystems, Courtaboeuf, France) with a de-clustering potential of 70 V. The molecular weight of the glucosylated products was confirmed by the presence of sodium adducts [m/z = M (substrate) + 180 (glucose) − 18 (H_2O) + 23 (sodium)] in positive mode.

Purification of glucosylation products was performed on a Waters Autopurif apparatus (Saint-Quentin-Fallavier, France) equipped with a 2545 pump, a 2996 photodiode array detector, a 3100 mass detector and a 2767 sample manager [Masslynx™ (Waters, Saint-Quentin-Fallavier, France) and Fractionlynx™ (Waters, Saint-Quentin-Fallavier, France) software]. A XBridge (Waters, Saint-Quentin-Fallavier, France) C18 column (4.6 × 150 mm) was used and the eluent solutions were 0.1% formic acid (eluent A) and acetonitrile with 0.1% formic acid (eluent B), using a 1.2 $mL{\cdot}min^{-1}$ elution rate and the gradient: 2% eluent B for 0.5 min then 2–16% eluent B over 0.5 min, 16-24% eluent B over 9 min. Double detection was done (both UV and MS detection). ^{1}H and ^{13}C-NMR spectra were obtained on Bruker, Wissembourg, France DPX300 or AV300 instruments using D_2O as the solvent.

Protein 3D modelling and ligand docking

The SlUGT5 protein homology model was prepared using Modeller 9.7 (with automodel default) [36], based on the UGT72B1 structure (PDB entry = 2VCE) (residues 6-476), after removal of all HETATM atoms and removing all alternative conformations (conformation A was retained for all alternative residues: Arg81, Ser87, Arg109, Leu118, Thr280, Glu284, Glu334, Arg405, Glu444, Arg448, Ser461). Eight ligands (hydroquinone, salicyl alcohol, methyl salicylate, guaiacol, eugenol, benzyl alcohol, phenyl ethanol and 4-OH benzyl alcohol) were drawn using the JME molecular editor (http://www.molinspiration.com/jme/index.html), transferred to the PRODRG2 server (http://davapc1.bioch.dundee.ac.uk/prodrg/) [41], and modelled using default parameters. PDB files were saved for docking analyses.

Docking was performed using AutoDock 4.2 and AutoDockTools 1.5.4. [42]. UDP-glucose from UGT72B1 was directly transferred into the SlUGT5 model without modification. For docking, the SlUGT5 model with UDP-glucose was considered as rigid. The catalytic histidine (His17) was considered as a flexible residue with only one torsion bond (CB-CG). Ligands were prepared using AutoDockTools and default parameters for the number of torsion angles and anchor definition. Box size was 31 × 31 × 31 points, with 0.375 Å spacing, manually centred on the acceptor molecule of the UGT72B1 structure. The Lamarkian genetic algorithm was used with 50 GA-LS runs and a maximum energy evaluation of 2 500 000 (medium). Clustering of the 50 conformations was performed using a 1 Å rmsd tolerance.

Acknowledgements

We are grateful to Gisele Borderies and Saida Danoun (UMR Surfaces Cellulaires et Signalisation chez les Végétaux, CNRS-UPS, Toulouse, France) for help during the HPLC analyses and initial LC-MS analyses, Ricardo Ayub and Marcela Yada (Universidade Estadual de Ponta Grossa, Departamento de Fitotecnia e Fitossanidade, University of Brasil, Brazil) for help with protein activity assays, Chris Ford (University of Adelaide, Australia) for the protein purification protocol, and Mondher Bouzayen, Jean-Claude Pech, Corinne Audran-Delalande, Mohamed Zouine and Alain Latche (UMR Génomique et Biotechnologie des Fruits, INRA-INP/ENSAT, Toulouse, France) for their support. We are also grateful to Wilfried Schwab (Department of Biotechnology of Natural Products, Technical University, Munich, Germany) for the generous gift of the FaGT2 construct, and to the Genomic platform team at Toulouse Genopole, where the quantitative PCR analyses were performed. Collaboration between INRA-INP/ENSAT and Plant and Food Research was initiated through funding from the Dumont D'Urville NZ/France Science and Technology support programme, and the collaboration between O.D.C. and C.C. was funded by an Institut National Polytechnique de Toulouse – Bonus Qualité Recherche grant.

References

1 Baldwin EA, Scott JW, Shewmaker CK & Schuch W (2000) Flavor trivia and tomato aroma: biochemistry and possible mechanisms for control of important aroma components. *HortScience* **35**, 1013–1022.
2 Birtic S, Ginies C, Causse M, Renard CMGC & Page D (2009) Changes in volatiles and glycosides during fruit maturation of two contrasted tomato (*Solanum lycopersicum*) lines. *J Agric Food Chem* **57**, 591–598.
3 Ortiz-Serrano P & Gil JV (2007) Quantification of free and glycosidically bound volatiles in and effect of glycosidase addition on three tomato varieties. *J Agric Food Chem* **55**, 9170–9176.
4 Buttery RG, Takeoka G, Teranishi R & Ling LC (1990) Tomato aroma components: identification of glycoside hydrolysis volatiles. *J Agric Food Chem* **38**, 2050–2053.
5 Marlatt C, Ho C & Chien MJ (1992) Tomato: studies of aroma constituents bound as glycosides in tomato. *J Agric Food Chem* **40**, 249–252.
6 Bowles DJ, Isayenkova J, Lim EK & Poppenberger B (2005) Glycosyltransferases: managers of small molecules. *Curr Opin Plant Biol* **8**, 254–263.

7 Gachon CMM, Langlois-Meurinne M & Saindrenan P (2005) Plant secondary metabolism glycosyltransferases: the emerging functional analysis. *Trends Plant Sci* **10**, 542–549.
8 Li Y, Baldauf S, Lim EK & Bowles DJ (2001) Phylogenetic analysis of the UDP-glycosyltransferase multigene family of *Arabidopsis thaliana*. *J Biol Chem* **276**, 4338–4343.
9 Fleuriet A & Macheix JJ (1985) Tissue compartmentation of phenylpropanoid metabolism tomatoes during growth and maturation. *Phytochemistry* **24**, 929–932.
10 Hefner T, Arend J, Warzecha H, Siems K & Stockigt J (2002) Arbutin synthase, a novel member of the NRD1β glycosyltransferase family, is a unique multifunctional enzyme converting various natural products and xenobiotics. *Bioorg Med Chem* **10**, 1731–1741.
11 Jones PR, Moller BL & Hoj PB (1999) The UDP-glucose:*p*-hydroxymandelonitrile-O-glucosyltransferase that catalyzes the last step in synthesis of the cyanogenic glucoside dhurrin in *Sorghum bicolor*. Isolation, cloning, heterologous expression, and substrate specificity. *J Biol Chem* **274**, 35483–35491.
12 Song JT, Koo YJ, Park JB, Seo YJ, Cho YJ, Seo HS & Choi YD (2009) The expression patterns of AtBSMT1 and AtSAGT1 encoding a salicylic acid (SA) methyltransferase and a SA glucosyltransferase, respectively, in Arabidopsis plants with altered defense responses. *Mol Cells* **28**, 105–109.
13 Lim EK, Doucet CJ, Li Y, Elias L, Worrall D, Spencer SP, Ross J & Bowles DJ (2002) The activity of Arabidopsis glycosyltransferases toward salicylic acid, 4-hydroxybenzoic acid, and other benzoates. *J Biol Chem* **277**, 586–592.
14 Hansen KS, Kristensen C, Tattersall DB, Jones PR, Olsen CE, Bak S & Moller BL (2003) The *in vitro* substrate regiospecificity of recombinant UGT85B1, the cyanohydrin glucosyltransferase from *Sorghum bicolor*. *Phytochemistry* **64**, 143–151.
15 Fukuchi-Mizutani M, Okuhara H, Fukui Y, Nakao M, Katsumoto Y, Yonekura-Sakakibara K, Kusumi T, Hase T & Tanaka Y (2003) Biochemical and molecular characterization of a novel UDP-glucose:anthocyanin 3′-*O*-glucosyltransferase, a key enzyme for blue anthocyanin biosynthesis, from gentian. *Plant Physiol* **132**, 1652–1663.
16 Jugdé H, Nguy D, Moller I, Cooney JM & Atkinson RG (2008) Isolation and characterization of a novel glycosyltransferase that converts phloretin to phlorizin, a potent antioxidant in apple. *FEBS J* **275**, 3804–3814.
17 Lairson LL & Withers SG (2004) Mechanistic analogies amongst carbohydrate modifying enzymes. *Chem Commun* **20**, 2243–2248.
18 Lairson LL, Henrissat B, Davies GJ & Withers SG (2008) Glycosyltransferases: structures, functions, and mechanisms. *Annu Rev Biochem* **77**, 521–555.
19 Wang X (2009) Structure, mechanism and engineering of plant natural product glycosyltransferases. *FEBS Lett* **583**, 3303–3309.
20 Lunkenbein S, Bellido M, Aharoni A, Salentijn EM, Kaldenhoff R, Coiner HA, Muñoz-Blanco J & Schwab W (2006) Cinnamate metabolism in ripening fruit. Characterization of a UDP-glucose:cinnamate glucosyltransferase from strawberry. *Plant Physiol* **140**, 1047–1058.
21 Ross J, Li Y, Lim E & Bowles DJ (2001) Higher plant glycosyltransferases. *Genome Biol* **2**, 3004.1–3004.6.
22 Witte S, Moco S, Vervoort J, Matern U & Martens S (2009) Recombinant expression and functional characterisation of regiospecific flavonoid glucosyltransferases from *Hieracium pilosella* L. *Planta* **229**, 1135–1146.
23 Loutre C, Dixon DP, Brazier M, Slater M, Cole DJ & Edwards R (2003) Isolation of a glucosyltransferase from *Arabidopsis thaliana* active in the metabolism of the persistent pollutant 3,4-dichloroaniline. *Plant J* **34**, 485–493.
24 Brazier-Hicks M & Edwards R (2005) Functional importance of the family one glucosyltransferase UGT72B1 in the metabolism of xenobiotics in *Arabidopsis thaliana*. *Plant J* **42**, 556–566.
25 Brazier-Hicks M, Edwards LA & Edwards R (2007) Selection of plants for roles in phytoremediation: the importance of glucosylation. *Plant Biotech J* **5**, 627–635.
26 Brazier-Hicks M, Offen WA, Gershater MC, Revett TJ, Lim EK, Bowles DJ, Davies GJ & Edwards R (2007) Characterization and engineering of the bifunctional N- and O-glucosyltransferase involved in xenobiotic metabolism in plants. *Proc Natl Acad Sci USA* **104**, 20238–20243.
27 Lanot A, Hodge D, Jackson RG, George GL, Elias L, Lim EK, Vaistij FE & Bowles DJ (2006) The glucosyltransferase UGT72E2 is responsible for monolignol 4-*O*-glucoside production in *Arabidopsis thaliana*. *Plant J* **48**, 286–295.
28 Lanot A, Hodge D, Lim EK, Vaistij FE & Bowles DJ (2008) Redirection of the flux through the phenylpropanoid pathway by increased glucosylation of soluble intermediates. *Planta* **228**, 609–616.
29 Pang Y, Peel GJ, Sharma SB, Tang Y & Dixon RA (2008) A transcript profiling approach reveals an epicatechin-specific glucosyltransferase expressed in the seed coat of *Medicago truncatula*. *Proc Natl Acad Sci USA* **37**, 14210–14215.
30 Zanor MI, Rambla JL, Chaib J, Steppa A, Medina A, Granell A, Fernie AR & Causse M (2009) Metabolic characterization of loci affecting sensory attributes in tomato allows an assessment of the influence of the levels of primary metabolites and volatile organic contents. *J Exp Bot* **60**, 2139–2154.

31 Smart CD, Myers KL, Restrepo S, Martin GB & Fry WE (2003) Partial resistance of tomato to *Phytophthora infestans* is not dependent upon ethylene, jasmonic acid, or salicylic acid signaling pathways. *Mol Plant Microbe Interact* **16**, 141–148.
32 Shimoda K, Yamane SY, Hirakawa H, Ohta S & Hirata T (2002) Biotransformation of phenolic compounds by the cultured cells of *Catharanthus roseus*. *J Mol Catal B Enzym* **16**, 275–281.
33 Syahrani A, Widjaja I, Indrayanto G & Wilkins AL (1998) Glucosylation of salicyl alcohol by cell suspension cultures of *Solanum laciniatum*. *J Asian Nat Prod Res* **1**, 111–117.
34 Yoon SH, Fulton DB & Robyt JF (2004) Enzymatic synthesis of two salicin analogues by reaction of salicyl alcohol with *Bacillus macerans* cyclomaltodextrin glucanyltransferase and *Leuconostoc mesenteroides* B-742CB dextransucrase. *Carbohydr Res* **339**, 1517–1529.
35 Kino K, Shimizu Y, Kuratsu S & Kirimura K (2007) Enzymatic synthesis of α-anomer-selective D-glucosides using maltose phosphorylase. *Biosci Biotechnol Biochem* **71**, 1598–1600.
36 Sali A & Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* **234**, 779–815.
37 Tikunov YM, de Vos RCH, Gonzalez Paramas AM, Hall RD & Bovy AG (2010) A role for differential glycoconjugation in the emission of phenylpropanoid volatiles from tomato fruit discovered using a metabolic data fusion approach. *Plant Physiol* **152**, 55–70.
38 Jones B, Frasse P, Olmos E, Zegzouti H, Li ZG, Latche A, Pech JC & Bouzayen M (2002) Down-regulation of DR12, an auxin-response-factor homolog, in the tomato results in a pleiotropic phenotype including dark green and blotchy ripening fruit. *Plant J* **32**, 603–613.
39 Alba R, Payton P, Fei Z, McQuinn R, Debbie P, Martin GB, Tanksley SD & Giovannoni JJ (2005) Transcriptome and selected metabolite analyses reveal multiple points of ethylene control during tomato fruit development. *Plant Cell* **17**, 2954–2965.
40 Chervin C, Tira-umphon A, Terrier N, Zouine M, Severac D & Roustan JP (2008) Stimulation of the grape berry expansion by ethylene and effects on related gene transcripts, over the ripening phase. *Physiol Plant* **134**, 534–546.
41 Schuettelkopf AW & van Aalten DMF (2004) PRODRG - a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallographica D* **60**, 1355–1363.
42 Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS & Olson AJ (2009) Autodock4 and AutoDockTools4: automated docking with selective receptor flexiblity. *J Comput Chem* **30**, 2785–2791.

Supporting information

The following supplementary material is available:
Fig. S1. Nucleotide sequences of the four *Solanum lycopersicum* UGT sequences (ORFs) cloned in this study.
Fig. S2. PAGE analysis of the soluble fractions of four recombinant SlUGT proteins.
Fig. S3. Sequence alignment of SlUGT5 with UGT homologues in group E.
Fig. S4. HPLC-UV traces of SlUGT5 glycosylation products.
Fig. S5. Structures of the four monoglucosides of salicyl alcohol.
Fig. S6. Surface representations of glycosyl transferase binding cavities.
Table S1. Subset of SlUGT5 (SGN-U315028) expression profiling data in the Tomato Functional Genomics Database.
Table S2. Approximated free binding energies and kI values for the binding clusters shown in Fig. 3.
Table S3. Approximated free binding energies and kI values for the binding clusters shown in Fig. 4.

This supplementary material can be found in the online version of this article.

Please note: As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.