Phytochemistry 70 (2009) 473–482

Contents lists available at ScienceDirect

Phytochemistry

journal homepage: www.elsevier.com/locate/phytochem

Substrate specificities of family 1 UGTs gained by domain swapping

Esben Halkjær Hansen [a,*], Sarah A. Osmani [b,c], Charlotte Kristensen [a], Birger Lindberg Møller [b,c], Jørgen Hansen [a]

[a] Evolva A/S, Bülowsvej 25, DK-1870 Frederiksberg C, Denmark
[b] Plant Biochemistry Laboratory, Department of Plant Biology and Biotechnology, University of Copenhagen, Thorvaldsensvej 40, DK-1871 Frederiksberg C, Copenhagen, Denmark
[c] VKR Research Centre "Pro-Active Plants", Thorvaldsensvej 40, DK-1871 Frederiksberg C, Copenhagen, Denmark

ARTICLE INFO

Article history:
Received 22 October 2008
Received in revised form 21 January 2009
Available online 2 March 2009

Keywords:
Domain swapping
Glycosylation
Glycosyltransferase
Biocatalysis
Resveratrol
Etoposide
Curcumin
ADME improvement
Combinatorial biochemistry

ABSTRACT

Family 1 glycosyltransferases are a group of enzymes known to embrace a large range of different substrates. This study devises a method to enhance the range of substrates even further by combining domains from different glycosyltransferases to gain improved substrate specificity and catalytic efficiency. Chimeric glycosyltransferases were made by combining domains from seven different family 1 glycosyltransferases, UGT71C1, UGT71C2, UGT71E1, UGT85C1, UGT85B1, UGT88B1 and UGT94B1. Twenty different chimeric glycosyltransferases were formed of which twelve were shown to be catalytically active. The chimeric enzymes of *Arabidopsis thaliana* UGT71C1 and UGT71C2 showed major changes in acceptor substrate specificity and were able to glycosylate etoposide significantly better than the parental UGT71C1 and UGT71C2 enzymes, with K_{cat} and efficiency coefficients 3.0 and 2.6 times higher, respectively. Chimeric glycosyltransferases of UGT71C1 combined with *Stevia rebaudiana* UGT71E1, also afforded enzymes with high catalytic efficiency, even though the two enzymes only display 38% amino acid sequence identity. These chimeras show a significantly altered regiospecificity towards especially *trans*-resveratrol, enabling the production of *trans*-resveratrol-β-4′-*O*-glucoside (resveratroloside). The study demonstrates that it is possible to obtain improved catalytic properties by combining domains from both closely as well as more distantly related glycosyltransferases. The substrate specificity gained by the chimeras is difficult to predict because factors determining the acceptor specificity reside in the N- terminal as well as the C-terminal domains.

1. Introduction

Plants synthesize a plethora of bioactive natural products (secondary metabolites) many of which are glycosides. Glycosylation serves to stabilize compounds, increase their solubility and to facilitate transport and storage. Plant glycosides classified as phytoanticipins are converted into their bioactive aglycons by specific β-glucosidases offering a swift chemical response towards herbivores (Morant et al., 2008). Glycoside formation is mediated by glycosyltransferases (GTs). The GTs constitute a large class of proteins that are divided into 91 families according to the CAZY classification system (http://www.cazy.org/fam/acc_GT.html) (Campbell et al., 1997; Coutinho et al., 2003; Osmani et al., 2009).The plant GTs that have small molecules as substrates belong to Family 1 and are characterized by their dependence on UDP-sugar donors for the glycosylation of bioactive natural products, hormones and xenobiotics, and are therefore termed UDP-sugar glycosyltransferases, or UGTs. The number of UGTs in a single plant species is high, e.g. 112 in *Arabidopsis thaliana* (Paquette et al., 2003) and 202 in *Oryza sativa* (http://www.cazy.org/geno/39947.html), underscoring their diverse and important biological roles in the synthesis of plant natural products. While some plant UGTs possess a quite broad acceptor specificity (Hefner et al., 2002; Vogt and Jones, 2000; Hansen et al., 2003; Kramer et al., 2003), others are reported as being highly specific, only glycosylating one or a few different acceptors (Fukuchi-Mizutani et al., 2003; Sawada et al., 2005; Kramer et al., 2003). UGT enzymes are divided into families based on their sequence identity. However, enzymes showing high amino sequence identity may possess very different substrate specificities (Weis et al., 2006) and substrate recognition often seems to be regio- rather than substrate specific (Jones et al., 1999; Vogt and Jones, 2000; Lim et al., 2003, 2004; Modolo et al., 2007). Thus, assignment of a UGT to a specific gene family does not necessarily provide unambiguous information about its substrate specificity. In addition to their *in planta* function, the high substrate diversity of plant UGTs can be utilized in the production of compounds that are difficult and costly to synthesize chemically. This has important applications in the medical industry, since glycosylation may improve absorption, distribution, metabolism as well as excretion (ADME properties) of drug molecules (Thorson et al., 2004).

* Corresponding author. Tel.: +45 35200235; fax: +45 35200231.
E-mail address: esbenh@evolva.com (E.H. Hansen).

0031-9422/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.phytochem.2009.01.013

The primary structure of the GTs in general shows low homology. Nevertheless, their secondary and tertiary structures are highly conserved (Unligil and Rini, 2000; Hu and Walker, 2002), as the GTs crystallized so far adopt one of three folds, named the GT-A, GT-B and GT-A-like folds (http://www.cermav.cnrs.fr/glyco3d/) (Osmani et al., 2009). The 3D fold appears to be conserved within each CAZY enzyme family (Hu and Walker, 2002; Zhang et al., 2003), and the Family 1 GTs, including the plant UGTs, adopt the GT-B fold. In plant UGTs, the GT-B fold comprises two domains formed by the N- and C-terminal parts, where the C-terminal harbors the PSPG motif which spans 44 amino acids and constitutes the only highly conserved amino acid sequence in the UGTs (Mulichak et al., 2001; Hughes and Hughes, 1994). The C-terminal domain and especially residues within the PSPG motif are responsible for the majority of interactions with the sugar donor, whereas the acceptor mainly interacts with residues harbored within the N-terminal domain (Mackenzie, 1990; Meech and Mackenzie, 1997; Ross et al., 2001; Lim et al., 2003; Osmani et al., 2009).

Mutational studies of a small number of plant UGTs have identified specific amino acid residues involved in substrate recognition, with some mutations resulting in UGTs with altered substrate specificity (Hefner and Stöckigt, 2003; Kubo et al., 2004; Modolo et al., 2007; Osmani et al., 2009). Homology modeling of plant UGTs based on crystal structures of bacterial Family 1 GTs has guided targeted mutagenesis of residues involved in sugar donor and acceptor binding (Hans et al., 2004; Thorsøe et al., 2005), and exchange of larger parts of primary sequence between similar UGTs without loss of functionality has also been demonstrated (Mackenzie, 1990; Masada et al., 2007; Weis et al., 2008; Cartwright et al., 2008). Within the last three years, four plant UGTs have been crystallized, revealing detailed information on substrate interactions and adding confidence to previous structures generated by homology modeling (Shao et al., 2005; Offen et al., 2006; Li et al., 2007; Osmani et al., 2008, 2009). The conservation of secondary and tertiary structure among the Family 1 plant UGTs and the apparent separation of donor and acceptor interacting domains supports the feasibility of a domain swapping approach to construct UGTs with novel catalytic specificities. Exchanging domains or parts of these with minimal impact on tertiary structure would enhance the possibility of forming functionally active chimera.

In the present paper we show that molecular homology modeling can be used for the rational design of such chimeras, even in the absence of exact crystal structures of the proteins in question. Our study was designed to thoroughly assess the effects of exchange of domains from both closely and more distantly related family 1 UGTs using domains defined by molecular models of the parental proteins. A selection of twenty chimeric proteins was constructed and we demonstrate that it is possible to obtain novel enzymes with improved catalytic properties as well as entirely new substrate specificities.

2. Results and discussion

2.1. Functionally active UGT chimeras between closely as well as distantly related enzymes

We set out to test the feasibility of domain swapping between closely as well as distantly related UGTs. The selected set of seven UGTs listed in Table 1 fulfils this criterion and covers a suitable range of different substrate specificities. The design of UGT chimeras was based on amino acid sequence alignments guided by predicted secondary structures as well as molecular modeling of the parent proteins. Based on the performed alignments, the chimeric UGTs were designed to position the assembly sites between elements of well conserved secondary and tertiary structural parts to obtain the least possible interference with 3D folding of the chimeric protein. The highly conserved 44 aa PSPG motif situated in the C-terminal region was in all swaps provided entirely by the C-terminal parental UGT. One set of chimeras was constructed with the assembly point situated between the main N- and C- terminal domains constituting the two distinct halves of the enzymes (Fig. 1A), and therefore envisioned as an optimal assembly point. A second set of chimeric UGTs were built using alternative assembly sites positioned in the N-terminal region (Fig. 1B and C). Finally, two chimeric UGTs composed of three assembled domains were constructed. This design strategy was chosen primarily to assess the possibilities of swapping various domains in UGTs, while retaining catalytic activity. In addition, this approach was expected to provide new information on the functional importance of particular domains in the constructed chimeric enzymes. The N-terminal domain constituted the main focus of our work, because this domain harbors the main part of acceptor substrate interactions.

Table 1
The series of UDP-glycosyltransferases selected for domain swapping experiments. References to previously reported substrate specificities are provided.

Enzyme	Species	Substrates	Reference
UGT71C1	*Arabidopsis thaliana*	Various benzoates and salicylic acid	Lim et al. (2002)
UGT71C2	*Arabidopsis thaliana*		Lim et al. (2003)
UGT71E1	*Stevia rebaudiana*		Richman et al. (2005)
UGT88B1	*Stevia rebaudiana*		Richman et al. (2005)
UGT85C1	*Stevia rebaudiana*	Terpenoid glycosyltransferase	Richman et al. (2005)
UGT85B1	*Sorghum bicolor*	Cyanohydrin glucosyltransferase	Hansen et al. (2003)
UGT94B1	*Bellis perennis*	Anthocyanin glucoronosyltransferase	Sawada et al. (2005) and Osmani et al. (2009)

The assembly sites were constructed by introduction of restriction sites at the fusion points. Preferentially, this was carried out by alteration of single nucleotides to introduce the restriction site without concomitant alteration of the encoded amino acid sequence. In those cases where this was not possible, the mutations were designed to result in conservative amino acid substitutions and preferably by introducing amino acids that according to the alignments made were present at the particular site in other UGTs. The fusions were made at four different sites, using *Aat*II, *Eco*RI, *Xma*I and *Apa*I restriction sites. Using the UGT71C1 amino acid numbers as a reference, *Aat*II were introduced at ^{77}Asn, *Eco*RI at ^{144}Phe, *Xma*I at ^{188}Pro and *Apa*I at ^{255}Gly. Molecular models were made of some of the designed chimeric UGTs to verify that the planned assembly sites did not interfere with important catalytical parts of the enzymes (Fig. 1).

A total of twenty chimeric UGTs were constructed (Fig. 2). Successful high level expression of six of the chimeras (up to 6 mg/l) was obtained in *Escherichia coli*. The six high yield proteins obtained were chimeras between UGT71C1 and UGT71C2, between UGT71E1 and UGT71C1 and between UGT71E1 and UGT71C2. Catalytic activity towards one or more sugar acceptors was monitored by HPLC analysis (Table 2). The yield from the remaining fourteen chimeras could not be determined accurately because the Coomassie stained band representing the UGT could not be unambiguously identified, due to very low concentration of soluble enzyme. In cases examined, this low recovery was due to the formation of inclusion bodies. Attempts to re-solubilize and reconstitute functional activity of UGTs deposited in inclusion bodies were unsuccessful. Nevertheless, four additional chimeras between UGT85C1

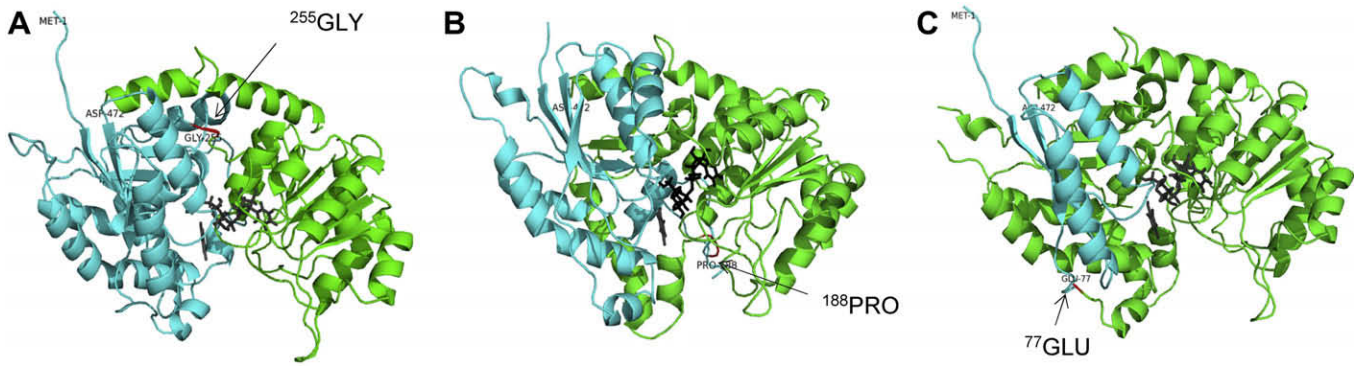

Fig. 1. Ribbon diagram 3D structures of the three UGT chimeras constructed from UGT71C1 and 71C2 by homology modeling UGT71C1$_{255}$71C2 (A), UGT71C1$_{188}$71C2 (B), UGT71C1$_{77}$71C2 (C). The fusion points of the constructed chimeras are shown. The N- terminal amino acids originating from 71C1 are colored in cyan and the C-terminal amino acids from 71C2 are shown in green. Amino acids at the fusion points are shown in red (see also Fig. 2). The sugar donor (UDP-glucose) and acceptor (2,4,5-trichlorophenol) are shown in black stick models. The assembly points between the domains were selected to minimize interference with well conserved secondary and tertiary structural parts of the enzymes.

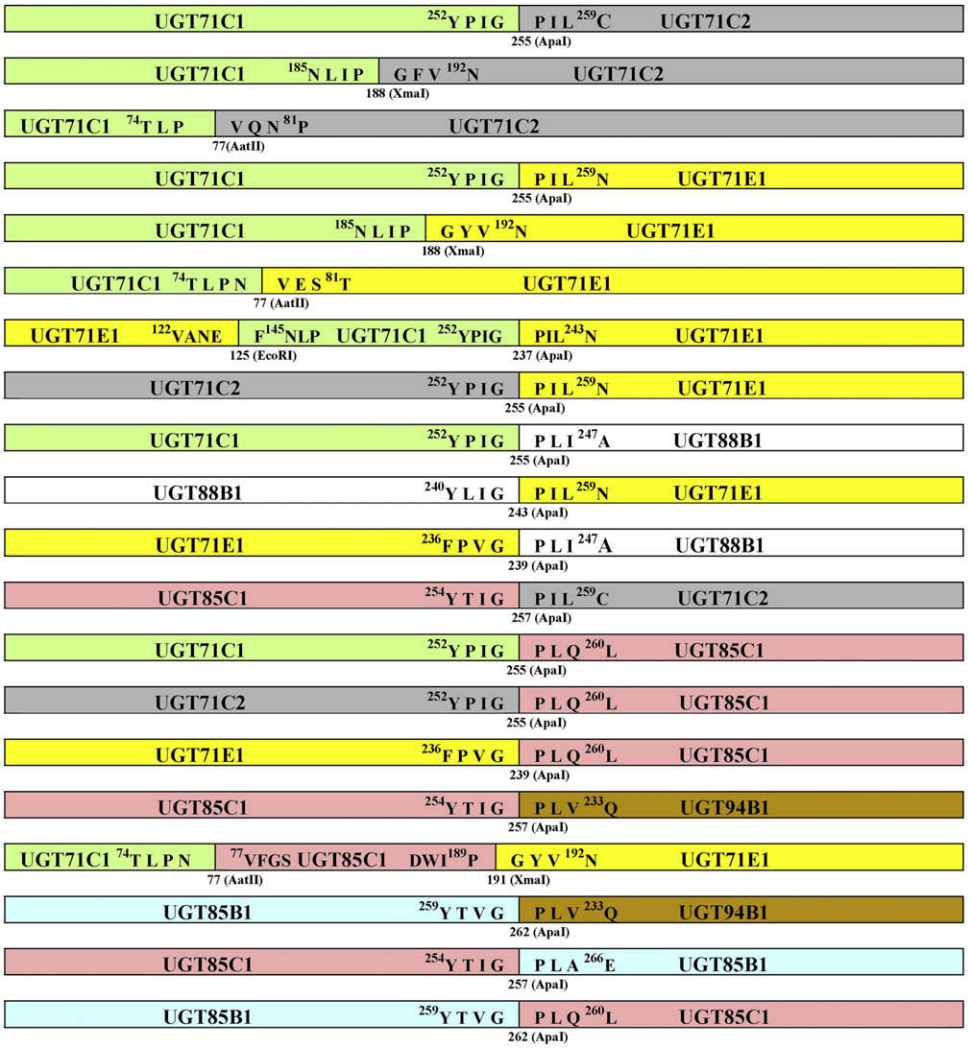

Fig. 2. Schematic overview of the composition of the twenty chimeric plant UGTs that were designed, produced as recombinant proteins in *E. coli* and purified by affinity chromatography. Amino acid numbers shown below the sites of assembly denotes the position of the amino acid residue in the chimeric enzyme which harbors the restriction site. The type of restriction site introduced is listed in parenthesis. The amino acid sequences and numbers at specific amino acids refer to the parental UGTs and serve to document the amino acid sequences around the site of assembly.

and UGT71C1, C2 or E1 were shown to be catalytically active in assays using [^{14}C]UDP-glucose as sugar donor. In control experiments with protein extracts obtained from *E. coli* cells harboring the empty pet30a+ expression vector, no background glycosylation of the substrates tested was detected neither using the HPLC detection method nor using radiolabelled [^{14}C]UDP-glucose. This

Table 2
The active UGTs, the enzyme yield of these and their ability to glycosylate various substrates.

Enzyme	Amino acid identity among parental UGTs	Enzyme yield	Glycosylated substrates
UGT71C1	–	6.0 mg l^{-1}	TCP, eugenol, *trans*-resveratrol
UGT71C2	–	1.5 mg l^{-1}	TCP, eugenol, curcumin
UGT71E1	–	2.2 mg l^{-1}	TCP, eugenol, curcumin, *trans*-resveratrol, etoposide
UGT88B1	–	1.5 mg l^{-1}	*Trans*-resveratrol
UGT85C1	–	1.0 mg l^{-1}	
UGT94B1	–	5.5 mg l^{-1}	Cyanidin 3-*O*-glucoside
UGT71C1$_{255}$71C2	78%	6.0 mg l^{-1}	TCP, eugenol, curcumin, *trans*-resveratrol, etoposide
UGT71C1$_{188}$71C2	78%	4.5 mg l^{-1}	TCP, eugenol, curcumin, *trans*-resveratrol, etoposide
UGT71C1$_{77}$71C2	78%	2.4 mg l^{-1}	TCP, eugenol, curcumin, *trans*-resveratrol, etoposide
UGT71C1$_{255}$71E1	38%	3.0 mg l^{-1}	TCP, eugenol, curcumin, *trans*-resveratrol, etoposide
UGT71E1$_{125}$-71C1$_{237}$71E1	38%	2.3 mg l^{-1}	TCP, eugenol, curcumin, *trans*-resveratrol
UGT71C2$_{255}$71E1	39%	2.3 mg l^{-1}	TCP, eugenol, curcumin, *trans*-resveratrol, etoposide
UGT85C1$_{257}$71C2	23%	<0.1 mg l^{-1}	[^{14}C] assay: detectable activity towards TCP and *trans*-resveratrol
UGT71C1$_{255}$85C1	22%	<0.1 mg l^{-1}	[^{14}C] assay: detectable activity towards TCP and *trans*-resveratrol
UGT71C2$_{255}$85C1	23%	<0.1 mg l^{-1}	[^{14}C] assay: detectable activity towards TCP and *trans*-resveratrol
UGT71E1$_{239}$85C1	22%	<0.1 mg l^{-1}	[^{14}C] assay: detectable activity towards TCP and *trans*-resveratrol

demonstrates that it is possible to construct functionally active chimeras between closely related sub-family members (e.g. UGT71C1 and UGT71C2, sharing 78% amino acid sequence identity), between members of different sub-families (e.g. UGT71E1 and UGT71C1, sharing 38% identity) as well as between enzymes of distantly related gene families (e.g. UGT85C1 and UGT71C1, sharing only 22% identity).

The kinetic properties of the six high yield chimeras between *A. thaliana* UGT71C1 or 71C2 and *Stevia rebaudiana* UGT71E1 were determined with the four substrates, vanillin, curcumin, etoposide and *trans*-resveratrol using UDP-glucose as sugar donor (Table 3, Fig. 3). The UDP-glucose was supplied at a concentration of 20 mM to ensure that saturable sugar donor conditions were administered to all tested enzymes. UDP-glucose kinetics experiments using saturating vanillin as acceptor substrate ensured, that this assumption was valid, that sugar kinetics followed normal Michaelis–Menten kinetics and that no substrate inhibition occurred at the 20 mM UDP-glucose as shown for the 71C1-71C2 chimeras incubated with 4 mM vanillin (Fig. 4).

2.2. New substrate specificities of UGT71C1-C2 chimeras

The three chimeras between UGT71C1 and UGT71C2 (71C1$_{255}$71C2, 71C1$_{188}$71C2 and 71C1$_{77}$71C2) showed altered acceptor substrate specificities. All three chimeras were able to glycosylate etoposide significantly better than either of the parental enzymes. The most effective chimera was UGT71C1$_{255}$71C2 for which the V_{max} and efficiency coefficient (E_f) were 3.0 and 2.6 times higher, respectively, than those of UGT71C1, the most active of the two parental enzymes (Table 3). The very low solubility of etoposide prevented the use of saturating substrate concentrations. This precluded accurate determination of absolute V_{max} values using the Michaelis–Menten equation. However, the etoposide glycosylation rate of the chimeric 71C1$_{255}$71C2 was 2–4 times higher than that of UGT71C1 at all substrate concentrations tested (Fig. 3), strongly indicating that the absolute V_{max} for the chimera has increased in comparison to the parental enzyme. Etoposide is derived from podophyllotoxin, a toxin found in American Mayapple, and is used as a topoisomerase II type inhibitor used for treatment of cancer malignances. The possibility to efficiently glucosylate etoposide thus has clear medical relevance with respect to improvement of its ADME characteristics.

The kinetic parameters for the three UGT71C1 and UGT71C2 based chimeras and for the two parental enzymes were also determined for the acceptor substrates curcumin, *trans*-resveratrol and vanillin. With vanillin as substrate, UGT71C1 was by far the most active enzyme while UGT71C2 was the most active enzyme using curcumin as substrate. With each of these two substrates, the chimeras show activities somewhat similar to the least efficient of the parental enzymes. A different situation is observed using *trans*-resveratrol as sugar acceptor. With this substrate, UGT71C1 is the most active and the UGT71C2 the least whereas the catalytic activities of the three chimeras are positioned in between as observed using vanillin and curcumin as sugar acceptors. However, when the concentration of *trans*-resveratrol was increased above 2 mM, the parental enzyme UGT71C1 showed a steady decrease in product formation when the substrate concentration was increased. The opposite effect was observed with the chimera 71C1$_{77}$71C2 which at the two highest substrate concentrations showed a product formation that exceeded that of UGT71C1. The significantly different results obtained with the different sugar acceptors support the notion that the increased activity of the chimeras towards etoposide is not due to a general increase in activity, e.g. by enhanced affinity towards UDP-glucose, but indeed to a profound alteration of the substrate specificity.

Several studies have suggested that the N-terminal and C-terminal domains of Family 1 UGTs may be considered as independent domains responsible for sugar acceptor and sugar donor binding, respectively, (Mackenzie, 1990; Meech and Mackenzie, 1997; Ross et al., 2001; Lim et al., 2003). This concept is not in agreement with the catalytic activities of the chimeric UGTs towards different sugar acceptors presented in this study, which show effects ascribed to the N- as well as C-terminal domains in a somewhat unpredictable manner. Several of the chimeras examined testify that the C-terminal part is important in determining affinity towards the sugar acceptor. The two chimeras 71C1$_{255}$71E1 and 71C2$_{255}$71E1 showed significantly higher affinities for etoposide than did either UGT71C1 or UGT71C2, a characteristic which must therefore be conferred by the C-terminal part of UGT71E1. Another example was obtained using curcumin as sugar acceptor. In this case, the 71C1$_{255}$71E1 chimera had at least three times higher activity compared to the UGT71C1 parental enzyme. The variable and somewhat unpredictable effects on the catalytic activity towards the sugar acceptors observed upon combination of N- and C- terminal domains from different UGTs, implies that the sugar acceptor specificity is a product of substrate interactions with both domains. These findings are in line with the study by He et al. (2006), who propose that Ala^{380}, which is a part of the C-terminal end of UGT71G1, affects the positioning of the acceptor substrate and with recent detailed studies documenting additional interactions with C-terminal residues (Osmani et al., 2008, 2009).

The possibility to swap domains among Family 1 glycosyltransferases has been the subject of several studies. A chimera constructed between the two *A. thaliana* UGTs 71C1 and 71C3 was shown to inherit the affinity towards UDP-glucose defined by the C-terminal UGT71C3 parental enzyme (Weis et al., 2008). This chimeric UGT lost 90% of the activity towards the tested acceptor sub-

Table 3
Kinetic parameters for UGT71 wild type and chimeric enzymes. Chimeric UGTs where catalytic activity has been improved compared to the parental enzymes are highlighted in bold.

	K_m (mM)	V_{max} (μmol min^{-1} mg^{-1})	K_{cat} (min^{-1})	E_f (1 mM substrate)
Curcumin				
UGT71C1	0.15 ± 0.01	0.022 ± 0.002	1.2 ± 0.1	6.1E−05 ± 0.5E−05
UGT71C2	0.59 ± 0.02	0.074 ± 0.002	3.9 ± 0.1	2.2E−04 ± 0.9E−05
UGT71C1$_{255}$71C2	0.15 ± 0.02	0.007 ± 0.001	0.37 ± 0.05	2.0E−05 ± 0.3E−05
UGT71C1$_{188}$71C2	0.07 ± 0.01	0.008 ± 0.001	0.42 ± 0.03	2.3E−05 ± 0.2E−05
UGT71C1$_{77}$71C2	0.11 ± 0.01	0.017 ± 0.001	0.89 ± 0.05	4.8E−05 ± 0.3E−05
UGT71E1	[b]0.002 ± 0.0001	[b]0.097 ± 0.005	[b]5.1 ± 0.3	[b]6.1E−04 ± 0.3E−04
UGT71C1$_{255}$71E1	0.19 ± 0.02	0.070 ± 0.005	3.9 ± 0.3	3.4E−04 ± 0.3E−04
UGT71C2$_{255}$71E1	0.07 ± 0.02	0.024 ± 0.004	1.3 ± 0.2	1.4E−04 ± 0.3E−04
UGT71E1$_{125}$71C1$_{237}$71E1	0.51 ± 0.03	0.129 ± 0.006	6.8 ± 0.3	4.1E−04 ± 0.3E−04
Vanillin				
UGT71C1	1.12 ± 0.04	1.198 ± 0.036	64.6 ± 2.0	1.8E−03 ± 0.1E−03
UGT71C2	1.18 ± 0.02	0.140 ± 0.002	7.4 ± 0.1	2.7E−04 ± 0.1E−04
UGT71C1$_{255}$71C2	0.64 ± 0.03	0.22 ± 0.01	11.6 ± 0.5	4.3E−04 ± 0.3E−04
UGT71C1$_{188}$71C2	0.36 ± 0.05	0.12 ± 0.01	6.3 ± 0.7	2.8E−04 ± 0.4E−04
UGT71C1$_{77}$71C2	0.63 ± 0.02	0.112 ± 0.004	5.9 ± 0.2	2.2E−04 ± 0.1E−04
UGT71E1	4.53 ± 0.03	3.2 ± 0.02	167.7 ± 1.2	2.0E−03 ± 0.03E−03
UGT71C1$_{255}$71E1	1.46 ± 0.02	0.403 ± 0.007	22.3 ± 0.1	6.8E−04 ± 0.1E−04
UGT71C2$_{255}$71E1	1.40 ± 0.02	0.154 ± 0.002	8.5 ± 0.1	2.7E−04 ± 0.1E−04
UGT71E1$_{125}$71C1$_{237}$71E1	2.16 ± 0.30	0.133 ± 0.016	7.1 ± 0.9	1.6E−04 ± 0.4E−04
Trans-resveratrol (Trans-resveratrol-3-O-glucoside formation)				
UGT71C1	[a]0.5 ± 0.1	0.063 ± 0.012	3.38 ± 0.65	8.0E−05 ± 2.0E−05
UGT71C2	[a]0.33 ± 0.02	0.010 ± 0.001	0.54 ± 0.03	2.4E−05 ± 0.2E−05
UGT71C1$_{255}$71C2	[a]0.31 ± 0.06	0.027 ± 0.004	1.42 ± 0.22	3.7E−05 ± 0.7E−05
UGT71C1$_{188}$71C2	[a]0.08 ± 0.01	0.028 ± 0.003	1.49 ± 0.14	4.3E−05 ± 0.4E−05
UGT71C1$_{77}$71C2	[a]0.32 ± 0.02	0.038 ± 0.002	2.02 ± 0.09	5.2E−05 ± 0.3E−05
UGT71E1	[b]0.8 ± 0.8	[b]0.6 ± 0.3	[b]34 ± 17	[b]1.1E−03 ± 1.4E−03
UGT71C1$_{255}$71E1	1.20 ± 0.05	0.064 ± 0.003	3.54 ± 0.14	9.7E−05 ± 0.6E−05
UGT71C2$_{255}$71E1	[a]0.16 ± 0.04	0.023 ± 0.005	1.25 ± 0.26	6.5E−05 ± 1.6E−05
UGT71E1$_{125}$71C1$_{237}$71E1	4.7 ± 0.5	0.009 ± 0.001	0.50 ± 0.05	5.3E−06 ± 1.1E−06
Trans-resveratrol (Trans-resveratrol-4′-O-glucoside formation)				
UGT71C1	0.8 ± 0.5	0.008 ± 0.003	0.5 ± 0.2	9.8E−06 ± 5.75E−06
UGT71C2	–	–	–	–
UGT71C1$_{255}$71C2	3.5 ± 0.2	0.023 ± 0.001	1.2 ± 0.05	1.4E−05 ± 0.1E−05
UGT71C1$_{188}$71C2	–	–	–	–
UGT71C1$_{77}$71C2	6.8 ± 2.0	0.002 ± 0.001	0.1 ± 0.03	8.1E−07 ± 3.89E−07
UGT71E1	[b]1.7 ± 0.2	[b]0.007 ± 0.001	[b]0.4 ± 0.03	[b]7.8E−06 ± 1.09E−06
UGT71C1$_{255}$71E1	5.7 ± 0.1	0.305 ± 0.005	16.9 ± 0.3	1.5E−04 ± 0.04E−06
UGT71C2$_{255}$71E1	8.1 ± 2.0	0.002 ± 0.000	0.11 ± 0.02	7.3E−07 ± 3.22E−07
UGT71E1$_{125}$71C1$_{237}$71E1	4.3 ± 0.5	0.13 ± 0.01	6.7 ± 0.7	7.6E−05 ± 1.42E−05
Etoposide				
UGT71C1	2.7 ± 0.2	0.013 ± 0.001	0.72 ± 0.06	1.2E−05 ± 0.2E−05
UGT71C2	31 ± 10	0.002 ± 0.000	0.09 ± 0.02	1.8E−07 ± 1.1E−07
UGT71C1$_{255}$71C2	3.1 ± 0.3	0.040 ± 0.004	2.1 ± 0.2	3.2E−05 ± 0.5E−05
UGT71C1$_{188}$71C2	1.3 ± 0.1	0.017 ± 0.002	0.9 ± 0.1	2.4E−05 ± 0.4E−05
UGT71C1$_{77}$71C2	1.9 ± 0.2	0.024 ± 0.002	1.3 ± 0.1	2.6E−05 ± 0.4E−05
UGT71E1	1.5 ± 0.1	0.17 ± 0.01	9.1 ± 0.5	2.7E−04 ± 0.3E−04
UGT71C1$_{255}$71E1	2.1 ± 0.2	0.070 ± 0.005	3.9 ± 0.3	8.9E−05 ± 1.2E−05
UGT71C2$_{255}$71E1	1.5 ± 0.1	0.067 ± 0.003	3.7 ± 0.2	1.1E−04 ± 0.1E−04
UGT71E1$_{125}$71C1$_{237}$71E1	6.1 ± 1.3	0.011 ± 0.002	0.6 ± 0.1	5.2E−06 ± 2.1E−06

[a] K_m values could not be accurately determined due to the lack of experimental data obtained at low substrate concentrations or due to a poor Michealis–Menten fit.
[b] In experiments with *trans*-resveratrol and curcumin as sugar acceptors and UGT71E1 as enzyme source, the glucosides formed serve as competing substrates for the enzyme. This results in overestimation of the K_m values and underestimation of the V_{max} values.

strate, scopoletin, in comparison to the parental enzymes (Weis et al., 2008). This chimeric UGT is very similar to the chimeric 71C1$_{255}$71C2 protein used in our current study, with the domain assembly points in the two constructs being only eight amino acids apart. The 5.5 and 3.2 fold decrease in turn-over kinetic that we obtained using vanillin and curcumin as sugar acceptors, respectively, thus corresponds well to the results obtained with scopoletin as substrate. However, when *trans*-resveratrol was tested as sugar acceptor with our chimeric UGT, the turn-over was only reduced by 28%, and was accompanied by a shift in regiospecificity resulting in glucosylation of *trans*-resveratrol at the 4′-*O* position instead of at the 3-*O* position (see below). As mentioned above, the 71C1$_{255}$71C2 chimeric UGT showed a turnover 2.8 fold higher with etoposide compared to the best parental enzyme. The results obtained with the various chimeras in our study would suggest that a more thorough testing of the 71C1-71C3 chimeric UGT constructed by Weis et al. (2008) with respect to additional sugar acceptors might identify substrates that were more effectively utilized by the chimeric UGT in comparison to its parental forms.

2.3. The chimeric enzymes UGT71C1$_{255}$71E1 and UGT71E1$_{125}$71C1$_{237}$71E1 show a shift in regiospecificity

The chimeric UGTs based on UGT71E1 and UGT71C1 or UGT71C2 as parental enzymes were also functionally active (Table 3). Some of the chimeras showed shifted regiospecificity towards *trans*-resveratrol. T*rans*-Resveratrol possesses two hydroxy groups that may be glycosylated to form either piceid (*trans*-resveratrol-3-*O*-β-glucoside, resveratroloside (*trans*-resveratrol-4′-*O*-β-glucoside) or *trans*-resveratrol-3,4′-di-*O*-β-glucoside. UGT71C1

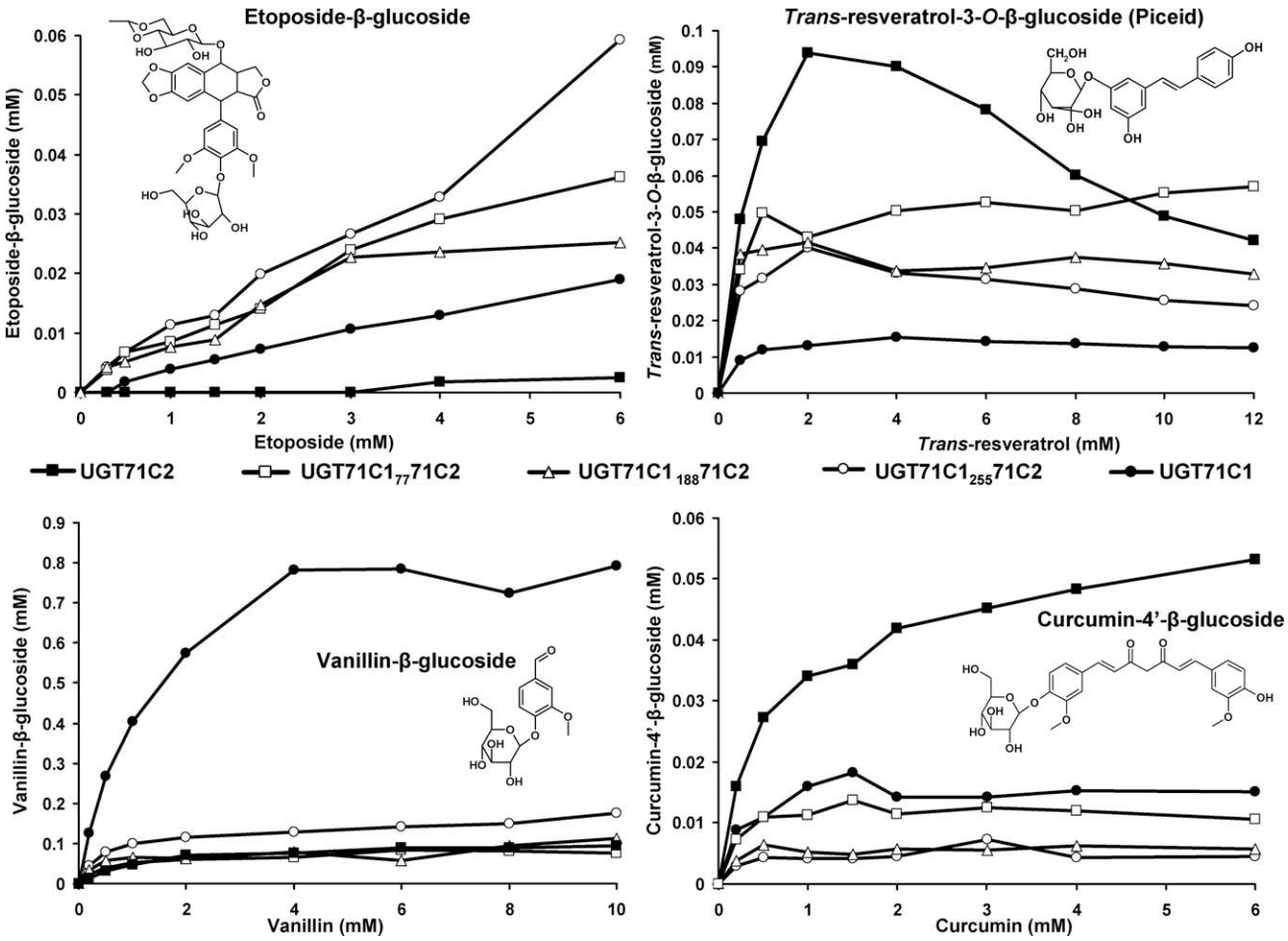

Fig. 3. Enzyme kinetic properties of the parental enzymes UGT71C1, UGT71C2 and of three derived chimeric enzymes using vanillin, curcumin, etoposide and *trans*-resveratrol as sugar acceptors and UDP-glucose as sugar donor. The structures of the mono β-glucosides formed are shown together with the saturation curves for the different sugar acceptors.

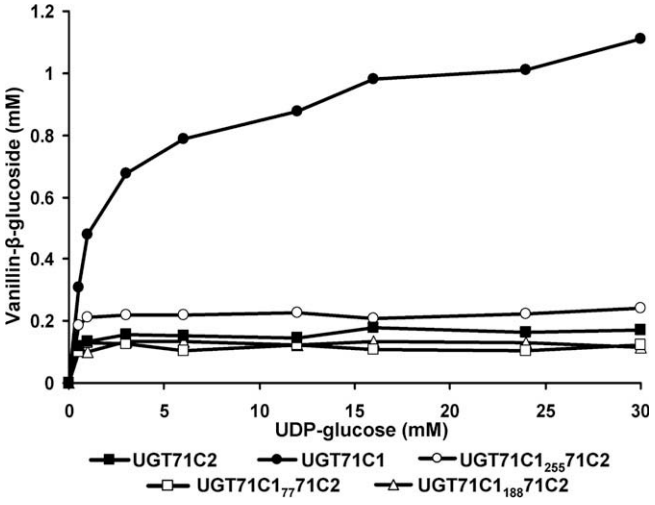

Fig. 4. Enzyme kinetic properties of the parental enzymes UGT71C1, UGT71C2 and of three derived chimeric enzymes, for UDP-glucose and 4 mM vanillin.

primarily catalyzed the formation of piceid, while UGT71E1 formed piceid and the diglucoside. Remarkably, the two chimeric UGT enzymes 71C1$_{255}$71E1 and 71E1$_{125}$71C1$_{237}$71E1 primarily gave rise to the formation of resveratroloside (Fig. 5). This demonstrates a shift in regiospecificity from the 3-*O* position of *trans*-resveratrol to the 4′-*O* position. As illustrated in Fig. 5, the turnover of the two chimeras was somewhat lower than for the highly active UGT71E1 parent, while being substantially higher than the 71C1 parent. The diglucoside *trans*-resveratrol-3,4′-di-*O*-β-glucoside formed by UGT71E1 could in theory be derived from resveratroloside, which could be turned over so fast that it was not observed to accumulate. If this situation applies, the formation of resveratroloside by the chimeric UGTs would be due to a loss rather than a gain of function. To resolve this ambiguity, a new set of experiments was carried out to test the ability of the parental enzyme UGT71E1 to use piceid and resveratroloside as substrates. Following an incubation period of 60 min 7.5% of the resveratroloside was glycosylated as compared to 99% of the piceid (data not shown). This showed that the UGT71E1 enzyme possesses a much higher affinity for piceid than resveratroloside and strongly indicate that the primary route to resveratrol-diglucoside synthesis by UGT71E1 proceeds *via* piceid as an intermediate. This demonstrates that the shift in regiospecificity of the chimeric UGTs 71C1$_{255}$71E1 and 71E1$_{125}$71C1$_{237}$71E1 towards the formation of resveratroloside indeed represents a gain of function.

The shifted regiospecificity observed for the chimeric UGT71E1$_{125}$71C1$_{237}$71E1 indicate that residues positioned within the exchanged amino acid region 125-239 of UGT71E1 are crucial for the regiospecificity of the enzyme. This observation correlates very well with the fact that several residues within this region have been observed to form part of the sugar acceptor pocket in the crystal based 3D structures of the plant UGTs *Vv*GT1, 71G1 and

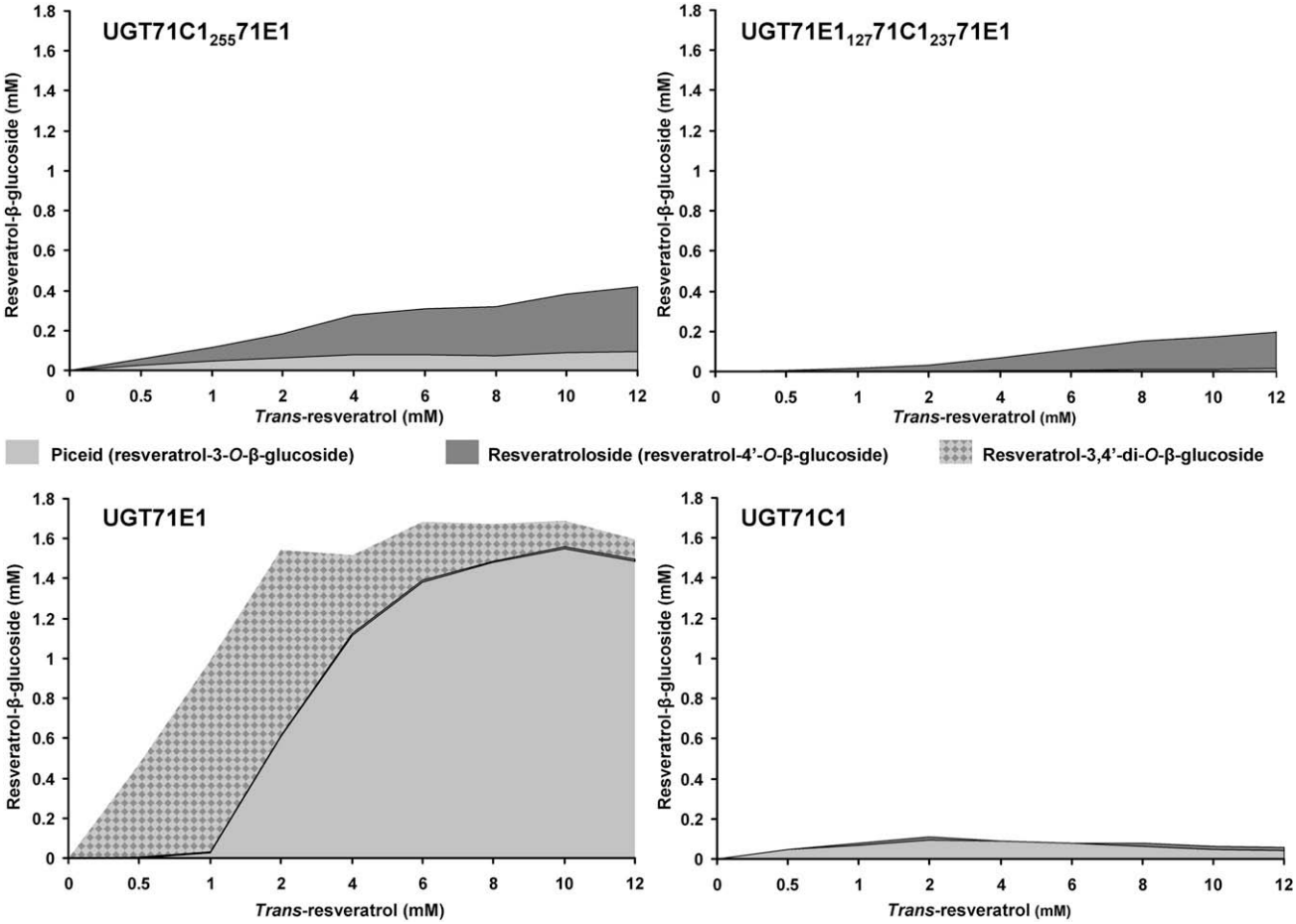

Fig. 5. The ability of the parental enzymes UGT71E1 and UGT71C1 and of two derived chimeric UGTs to form different types of *trans*-resveratrol glucosides. Resveratroloside (*trans*-resveratrol-4′-*O*-β-glucoside) is the major product formed by the chimeric enzymes UGT71C1$_{255}$71E1 and UGT71E1$_{125}$71C1$_{237}$71E1. Piceid (*trans*-resveratrol-3-*O*-β-glucoside) is the main product produced by the native parental UGT71E1 and UGT71C1, but UGT71E1 also produces the diglucoside *trans*-resveratrol-3,4′-di-*O*-β-glucoside.

72B1 (Shao et al., 2005; Offen et al., 2006; Brazier-Hicks et al., 2007). Specific amino acids that are directly involved in defining regiospecificity have also been identified within this region. This has been demonstrated in *Medicago truncatula* UGT71G1, where the mutations Y202A and F148 V shifted the regiospecificity with respect to glucosylation of quercetin and genistein (He et al., 2006). A domain swapping strategy using UGT74F1 and UGT74F2 (90% sequence identity) as parental enzymes resulted in the identification of a single amino acid mutation N142Y in UGT74F1 that shifted the regiospecificity of this UGT with respect to glucosylation of quercetin from the 7-*O*, 3′-*O* and 4′-*O* positions towards the 4′-*O* position (Cartwright et al., 2008). The analogous positions of these three regiospecificity shifting amino acids in UGT71E1 were difficult to define because this region in plant UGTs is one of the most diverging within the UGTs making exact alignment difficult (Osmani et al., 2009), but most likely ^{148}F and ^{202}Y in 71G1 corresponds to ^{141}L and ^{197}L UGT71E1 and ^{142} N of UGT74F1 corresponds to ^{144}M, all being points that could be targets for a mutational study of UGT71E1.

3. Experimental

3.1. Prediction of secondary protein structures and 3D homology modeling

Amino acid sequence alignments including predicted secondary structures (α-helices and β-strands) were made encompassing all the selected UGTs, using the program Phyre (Protein Homology/analogY Recognition Engine) (Kelley et al., 2000; Nett-Lovsey et al., 2008). 3D structures based on homology modeling were made using the Composer software Sybyl (Tripos, St. Louis MO, USA). Models were built using the crystal structure coordinates of *Mt*UGT72C1 (Shao et al., 2005) and *Vv*GT1 (Offen et al., 2006) as scaffolds.

3.2. Molecular biology techniques

Restriction digestion of DNA was performed according to the manufacturer (New England Biolabs, Ipswich MA, USA). Gel electrophoresis was performed in 1% agarose gels (Invitrogen, Carlsbad CA, USA) with 0.04 μl/ml ethidium bromide, Tris-Acetate EDTA buffer (TAE-buffer) as running buffer, and the 1 kb ladder plus® (Invitrogen) as DNA size marker. Ligation reactions (RT, 1-2 h) were carried out using T4-ligase (400.000 U/ml, 10% of reaction volume) and the supplied T4-ligase buffer (New England Biolabs). Small scale preparation of plasmid DNA from *E. coli* was generally performed using the alkaline lysis method (Sambrook and Russell, 2001). Plasmid DNA for cloning and sequencing purposes were prepared using the GeneJET Plasmid Miniprep kit (Fermentas).

3.3. Polymerase chain reactions

Polymerase chain reactions were performed in 200 μl polyethylene low profile thermo strips or tubes (ABgene) using a Peltier

thermal cycler DNA engine DYAD™ PCR machine. PCR grade dNTPs (Invitrogen) and *Pwo* polymerase (Roche Diagnostics, Hvidovre, Denmark) were used for all reactions. PCRs were performed with an initial preheating step at 94 °C for 2 min. and a final 4 min elongation step at the used elongation temperature. Thee oligonucleotides used to design the chimeric UGTs are displayed in Table 4.

3.4. Protein expression and purification

The chimeric UGT proteins were produced in *E.coli* XJb(DE3) Autolysis™ (*E.coli* B F- ompT $hsdS_B$ (r_B^- m_B^-) gal dcm^+ araB::R, cat λDE3) (Zymo research, Orange CA, USA) harboring the pet30a+ expression vector (Novagen, Nottingham, UK), which carries an N-terminal 6xHis tag sequence that was used for affinity purification. The cultures were grown in 1500 ml NZCYM (pH 7.0) containing 15 g Tryptone, 7.5 g NaCl, 7.5 g yeast extract, 1.5 g casamino acids, 3 g $MgSO_4$ and fortified with 30 mg/l kanamycin, 0.1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and 3 mM L-arabinose. After incubation (20 h, 20 °C), cells were pelleted and lysed in 25 ml lysis buffer (10 mM Tris–HCl (pH 7.5)), 5 mM $MgCl_2$, 1 mM $CaCl_2$, 3 tablets/100 ml Complete mini protease inhibitor cocktail (Roche diagnostics), 14 mg/l deoxyribonuclease (Calbiochem, Nottingham, UK) by a single freeze-thaw cycle to release lysozyme from cell cytoplasm. Purification was performed by adding 1/3 vol of 4x binding buffer (2 M NaCl, 80 mM Tris–HCl (pH 7.5)) to the lysate supernatant and followed by incubation (2 h) with His-select™ Nickel affinity gel (Sigma-Aldrich, Brøndby, Denmark). The affinity gel was recovered by centrifugation and the chimeric UGT eluted by adding elution buffer (7.5 ml 20 mM Tris–HCl (pH 7.5), 500 mM NaCl and 250 mM imidazole). Eluted protein was stabilized by addition of glycerol to a final concentration of 50%. SDS–PAGE was performed usingNuPAGE® 4–12% Bis–Tris 1.0 mm precast gels (Invitrogen), NuPAGE MOPS (Invitrogen) running buffer and Simplyblue Safestain (Invitrogen) for the Coomassie based gel staining. The amount of UGTs produced was determined semi quantitatively from the staining intensity of the observed UGT band using bovine serum albumin (Sigma-Aldrich) as reference. Fourteen of the UGTs were expressed in so low amounts that a specific Coomassie stained band could not be assigned to these UGTs. Thus the amount of each of these UGTs used in the enzymatic incubation mixtures was not determined.

3.5. Glycosylation assays using radiolabelled UDPG

Glycosylation reactions were performed in 96 well microtiter plates. Enzyme assays (total volume: 30 μl) contained 0.75 μg enzyme, 100 mM Tris–HCl (pH 8), 1 mM $MgCl_2$, 0.1 U/μl calf intestine phosphatase (Fermentas, Helsingborg, Sweden), 8.2 μM [^{14}C]UDP-glucose (7.4 GBq/mmol) GE healthcare, Hillerød, Denmark) and 2 mM acceptor substrate (dissolved in DMSO). After incubation (30 °C, 20 h), aliquots (5 μl) of the reaction mixture were applied to Silica gel 60 F_{254} TLC sheets (Merck, Glostrup, Denmark) and developed in EtOAc:Me_2CO:CH_2Cl_2:MeOH:H_2O (20:15:6:5:4 (vol/vol)). The TLCs were exposed (96 h) to phosphorimager screens (Amersham biosciences), and radiolabelled glucoside formed visualized and quantified using a STORM phosphoimager (Amersham biosciences).

3.6. Enzyme kinetics

Kinetic analyses of different acceptor substrates were performed using 20 mM UDP-glucose as sugar donor (Roche, Hvidovre, Denmark). The following acceptor substrates were tested: vanillin, etoposide, curcumin (Sigma-Aldrich), *trans*-resveratrol (Cayman chemicals, Tallinn, Estonia), and piceid (Alexis, San Diego, CA, USA). Resveratroloside was purified from an enzymatic glycosylation reaction (as described above), employing *At*UGT72B2 (NC_003070, nt:148319–149761) in a 25 ml glycosylation reaction mixture. Purification was performed using preparative HPLC using an Agilent 1200 series preparative HPLC system (Agilent Technologies, Nærum, Denmark) fitted with a Thermo Biobasic 18 column (150 × 30 mm, 10 μm particles, 150 Å pore size) (ThermoFisher Scientific, Waltham MA, USA).

Kinetic analyses of sugar donor efficiency were performed using 4 mM vanillin as sugar acceptor. Enzyme kinetic parameters were obtained by fitting the kinetics data to the Michaelis–Menten equation. At the experimental conditions chosen, all performed reactions were single-substrate saturable enzyme reactions as achieved by adding the second substrate in amounts saturating the enzyme. For some enzymes tested with *trans*-resveratrol, significant substrate inhibition was observed at resveratrol concentrations above 2 mM which reduced the number of data points to be fitted to Michealis–Menten equation. In such cases, V_{max} was defined as the highest observed conversion rate, while determination of K_m was done by Michealis–Menten fit. Etoposide has a low solubility which made it impossible to test at high substrate concentrations. Accordingly, all V_{max} values were defined as the highest observed conversion rate with a particular enzyme. For any practical comparison of activity between the parental enzymes and chimeras thereof, this experimental approach provides more valid data in comparison to extrapolated values resulting in unrealistic high kinetic parameters.

Table 4
Oligonucleotides used to design chimeric UGTs.

Primer	Sequence	Restriction site
71C1_front_F_NcoI	ATTA CCATGG CTATGGGGAAGCAAGAAGATGC	*Nco*I
71C1_front_R_ApaI	ATTA G GGCCCGATTGGGTAAATGGTT	*Apa*I
71C1_back_F_ApaI	ATTA GGGCCC ATATTATGCTCCAACGACCG	*Apa*I
71C1_back_R_NotI	ATTA GCGGCCGC CTACTTACTTATAGAAACGC	*Not*I
71C2_front_F_NcoI	ATTA CCATGG CTATGGCGAAGCAGCAAGAAGC	*Nco*I
71C2_front_R_ApaI	ATTA G GGCCCGATTG GGTAAACGGG	*Apa*I
71C2_back_F_ApaI	ATTA GGGCCC ATTCTATGCTCCAACGATCG	*Apa*I
71C2_back_R_NotI	ATTA GCGGCCGC TCAAAGCCCATCTATGAATC	*Not*I
71E1_front_F_PciI	ATTA ACATGT CTATGTCCACCTCAGAGCTTGTT	*Pci*I
71E1_front_R_ApaI	ATTA GGGC CCAACCGGGAAAACAGGTGG	*Apa*I
71E1_back_F_ApaI	ATTA GGGCCC ATTTTGAACCTTGAAAACAAAAAAG	*Apa*I
71E1_back_R_NotI	ATTA GCGGCCGC TTAAATCGTAACATTCGATA C	*Not*I
71C1_B2B_F_AatII	ATTA GACGTC CAAGACCCTCCACCAATG	*Aat*II
71C1_B4B_F_XmaI	ATTA CCCGGG TATGTCAACTCTGTTCCTAC	*Xma*I
71C2_B2back_F	ATTA GACGTC CAAAACCCTCCACCAATGGAG	*Aat*II
71C2_B4back_F	ATTA CCCGGG TTTGTTAACTCCGTTCCGG	*Xma*I
71C1_B1-B2F_R_AatII	ATTA GACG TC GGGCAACGTAACGAGACGG	*Aat*II
71C1_B1-B4F_R_XmaI	ATTA CCCG GG AATGAGATTCAAC TCCTC	*Xma*I
71E1_B2R_F_AatII	ATTA GACGTC GAGTCCACAATGGCTCTCATCTC	*Aat*II
71E1_B4R_F_XmaI	ATTA CCCGGG TATGTTAACCCGGTTCCTGC	*Xma*I
85C1_front_F_PciI	ATTA ACATGT CTATGGATCAAATGGCAAAAATT	*Pci*I
85C1_front_R_ApaI	ATTA G GGCCCAATGGTGTAAACATTAG	*Apa*I
85C1_back_F_ApaI	ATTA GGGCCC CTCCAGTTGCTTTTGAACAAAAT	*Apa*I
85C1_back_R_NotI	ATTA GCGGCCGC CTAGTTTCTTGATAACTTCT TG	*Not*I
85B1_front_F_PciI	ATTA ACATGT CTATGGGCAGCAACGCGCCGC	*Pci*I
85B1_front_R_ApaI	ATTA G GGCCCCACGGTGTAGATCGG	*Apa*I
85B1_back_F_ApaI	ATTA GGGCCC CTCGCCGAGGTCATCGCGTC	*Apa*I
85B1_back_R_NotI	ATTA GCGGCCGC TTACTGCTTGCCCCCGACCAG	*Not*I
94B1_front_F_PciI	ATTA ACATGT CTATGGATTCAAAAATCGATTC	*Pci*I
94B1_front_R_ApaI	ATTA GGGC CCAACCGGCA ATACTTTCTTTC	*Apa*I
94B1_back_F_ApaI	ATTA GGGCCC TTAGTTCAAGAAGCTTCTTTA	*Apa*I
94B1_back_R_NotI	ATTA GCGGCCGC TCAATTATTCATTTCACAAA G	*Not*I
85C1_B2_F_AatII	ATTA GACGTC TTCGGTTCTGCTAAAGACGACG	*Aat*II
85C1_B4_R_XmaI	ATTA CCCG GG AATCCAGTCTATTTCCATG	*Xma*I
71C1_B3_EF_EcoRI	ATTA GAATTC AATCTCCCTTCTTACATTTTC	*Eco*RI
71E1_B2F_R_AatII	ATTA GACGTC AGGGATGTCAACGAACCGTA	*Aat*II
71E1_B4F_R_XmaI	ATTA CCCGGG AACAGACAACTCAGTATCCG	*Xma*I
71E1_B3F_R_EcoRI	ATTA GAATTC GTTTGCAACATCACTCATCG	*Eco*RI

The efficiency function E_f was used as the main parameter for comparison of catalytic efficiency between the different enzyme species tested (Ceccarelli et al., 2008). The efficiency function is defined as: $E_f = k_{cat}/k_{dif}$ (K_m + [S]) ($K_{dif} = 10^9$ $M^{-1}s^{-1}$) and based on an irreversible reaction as experimentally obtained by removal of the UDP formed by the added phosphatase. The classic k_{cat}/K_m specificity constant was not used due to the unsuitability of the constant for describing differences between different enzyme species (Eisenthal et al., 2007).

3.7. LC-MS analysis

Enzyme catalyzed glucoside formation was analysed by LC-MS using an Agilent 1100 Series HPLC (Agilent Technologies) system fitted with a Hypersil gold C18 column (100 × 2.1 mm, 3 μm particles, 80 Å pore size) (ThermoFisher Scientific, Waltham MA, USA) and hyphenated to a TSQ Quantum (ThermoFisher Scientific) triple quadropole mass spectrometer with electrospray injection. As mobile phase (flow rate: 0.5 ml/min, 30 °C) was used a gradient of MeCN and H_2O adjusted to pH 2.3 with H_2SO_4. The gradient was composed as follows: 10% MeCN for 0.5 min, linear gradient from 10% to 100% for 6 min followed by 100% MeCN for 1 min. A diode array detector was used to monitor elution of aromatic compounds by UV-fluorescence. *Trans*-resveratrol was quantified based on its absorbance at 307 nm, curcumin at 425 nm, etoposide and vanillin at 230 nm. Glycosides formed were quantified at the same wavelength as their respective aglycons, assuming that glycoside and aglycon absorbs equally. This assumption was validated by comparing the amount of glycoside formed with the amount of aglycon that disappeared.

4. Conclusions

The use of plant family 1 UGTs for biological synthesis of glucosides of complex natural and non-natural aglycons provides an interesting route to compounds that otherwise are difficult and costly to obtain by chemical synthesis. From a commercial point of view, the process requires a glycosyltransferase that possesses high activity for the desired combination of sugar donor and sugar acceptor. Screening of a number of native enzymes obtained from different plant species may offer an enzyme with the desired properties. But this is not always the case. In this study we investigated the potential of domain swapping to create chimeric UGTs with an expanded catalytic range of sugar acceptors in comparison to the parental enzymes. A test of twenty different chimeric UGTs designed using domains from seven parental UGTs demonstrated the great potential of the approach in spite of the fact that only few sugar acceptors were examined as substrates. We obtained twelve active chimeric enzymes from domain swaps sharing as little as 22% sequence identity. Especially active chimeric UGTs were obtained upon combination of domains between *A. thaliana* UGT71C1, UGT71C2 and *S. rebaudiana* UGT71E1 and the improved overall activity was accompanied by significant changes in both substrate- and regiospecificity. In general, the approach to enlarge enzyme functionality through the production of chimeric enzymes turned out to prove highly feasible and the results obtained thereby serve to emphasize the possibilities of successful domain swapping among Family 1 UGTs.

References

Brazier-Hicks, M., Offen, W.A., Gershater, M.C., Revett, T.J., Lim, E.K., Bowles, D.J., Davies, G.J., 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.

Campbell, J.A., Davies, G.J., Bulone, V., Henrissat, B., 1997. A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities. Biochem. J. 326, 929–939.

Cartwright, A.M., Lim, E.K., Kleanthous, C., Bowles, D.J., 2008. A kinetic analysis of regiospecific glucosylation by two glycosyltransferases of *Arabidopsis thaliana*: domain swapping to introduce new activities. J. Biol. Chem. 283, 15724–15731.

Ceccarelli, E.A., Carrillo, N., Roveri, O.A., 2008. Efficiency function for comparing catalytic competence. Trends Biotechnol. 26, 117–118.

Coutinho, P.M., Deleury, E., Davies, G.J., Henrissat, B., 2003. An evolving hierarchical family classification for glycosyltransferases. J. Mol. Biol. 328, 307–317.

Eisenthal, R., Danson, M.J., Hough, D.W., 2007. Catalytic efficiency and k_{cat}/K_M: a useful comparator? Trends Biotechnol. 25, 247–249.

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.

Hans, J., Brandt, W., Vogt, T., 2004. Site-directed mutagenesis and protein 3D-homology modelling suggest a catalytic mechanism for UDP-glucose-dependent betanidin 5-*O*-glucosyltransferase from *Dorotheanthus bellidiformis*. Plant J. 39, 319–333.

Hansen, K.S., Kristensen, C., Tattersall, D.B., Jones, P.R., Olsen, C.E., Bak, S., Møller, B.L., 2003. The *in vitro* substrate regiospecificity of recombinant UGT85B1, the cyanohydrin glucosyltransferase from *Sorghum bicolor*. Phytochemistry 64, 143–151.

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

Hefner, T., Arend, J., Warzecha, H., Siems, K., Stöckigt, J., 2002. Arbutin synthase, a novel member of the NRD1beta glycosyltransferase family, is a unique multifunctional enzyme converting various natural products and xenobiotics. Bioorg. Med. Chem. 10, 1731–1741.

Hefner, T., Stöckigt, J., 2003. Probing suggested catalytic domains of glycosyltransferases by site-directed mutagenesis. Eur. J. Biochem. 270, 533–538.

Hu, Y., Walker, S., 2002. Remarkable structural similarities between diverse glycosyltransferases. Chem. Biol. 9, 1287–1296.

Hughes, J., Hughes, M.A., 1994. Multiple secondary plant product UDP-glucose glucosyltransferase genes expressed in cassava (*Manihot esculenta Crantz*) cotyledons. DNA Seq. 5, 41–49.

Jones, P.R., Møller, B.L., Høj, P.B., 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.

Kelley, L.A., MacCallum, R.M., Sternberg, M.J., 2000. Enhanced genome annotation using structural profiles in the program 3D-PSSM. J. Mol. Biol. 299, 499–520.

Kramer, C.M., Prata, R.T., Willits, M.G., De Luca, V., Steffens, J.C., Graser, G., 2003. Cloning and regiospecificity studies of two flavonoid glucosyltransferases from *Allium cepa*. Phytochemistry 64, 1069–1076.

Kubo, A., Arai, Y., Nagashima, S., Yoshikawa, T., 2004. Alteration of sugar donor specificities of plant glycosyltransferases by a single point mutation. Arch. Biochem. Biophys. 429, 198–203.

Li, L., Modolo, L.V., Escamilla-Trevino, L.L., Achnine, L., Dixon, R.A., Wang, X., 2007. Crystal structure of *Medicago truncatula* UGT85H2-insights into the structural basis of a multifunctional (iso)flavonoid glycosyltransferase. J. Mol. Biol. 370, 951–963.

Lim, E.K., Ashford, D.A., Hou, B., Jackson, R.G., Bowles, D.J., 2004. *Arabidopsis* glycosyltransferases as biocatalysts in fermentation for regioselective synthesis of diverse quercetin glucosides. Biotechnol. Bioeng. 87, 623–631.

Lim, E.K., Baldauf, S., Li, Y., Elias, L., Worrall, D., Spencer, S.P., Jackson, R.G., Taguchi, G., Ross, J., Bowles, D.J., 2003. Evolution of substrate recognition across a multigene family of glycosyltransferases in *Arabidopsis*. Glycobiology 13, 139–145.

Lim, E.K., Doucet, C.J., Li, Y., Elias, L., Worrall, D., Spencer, S.P., Ross, J., Bowles, D.J., 2002. The activity of *Arabidopsis* glycosyltransferases toward salicylic acid, 4-hydroxybenzoic acid, and other benzoates. J. Biol. Chem. 277, 586–592.

Mackenzie, P.I., 1990. Expression of chimeric cDNAs in cell culture defines a region of UDP glucuronosyltransferase involved in substrate selection. J. Biol. Chem. 265, 3432–3435.

Masada, S., Terasaka, K., Mizukami, H., 2007. A single amino acid in the PSPG-box plays an important role in the catalytic function of *Ca*UGT2 (Curcumin glucosyltransferase), a Group D Family 1 glucosyltransferase from *Catharanthus roseus*. FEBS Lett. 581, 2605–2610.

Meech, R., Mackenzie, P.I., 1997. Structure and function of uridine diphosphate glucuronosyltransferases. Clin. Exp. Pharmacol. Physiol. 24, 907–915.

Modolo, L.V., Blount, J.W., Achnine, L., Naoumkina, M.A., Wang, X., Dixon, R.A., 2007. A functional genomics approach to (iso)flavonoid glycosylation in the model legume *Medicago truncatula*. Plant Mol. Biol. 64, 499–518.

Morant, A.V., Jørgensen, K., Jørgensen, C., Paquette, S.M., Sánchez-Pérez, R., Møller, B.L., Bak, S., 2008. Beta-Glucosidases as detonators of plant chemical defense. Phytochemistry 69, 1795–1813.

Mulichak, A.M., Losey, H.C., Walsh, C.T., Garavito, R.M., 2001. Structure of the UDP-glucosyltransferase GtfB that modifies the heptapeptide aglycone in the biosynthesis of vancomycin group antibiotics. Structure 9, 547–557.

Nett-Lovsey, R.M., Herbert, A.D., Sternberg, M.J., Kelley, L.A., 2008. Exploring the extremes of sequence/structure space with ensemble fold recognition in the program Phyre. Proteins 70, 611–625.

Offen, W., Martinez-Fleites, C., Yang, M., Lim, E.K., Davis, B.G., Tarling, C.A., Ford, C.M., Bowles, D.J., Davies, G.J., 2006. Structure of a flavonoid glucosyltransferase reveals the basis for plant natural product modification. EMBO J. 25, 1396–1405.

Osmani, S.A., Bak, S., Imberty, A., Olsen, C.E., Møller, B.L., 2008. Catalytic key amino acids and UDP sugar donor specificity of a plant glucuronosyl transferase, UGT94B1. Molecular modeling substantiated by site-specific mutagenesis and biochemical analyses. Plant Physiol. 148, 1295–1308.

Osmani, S.A., Bak, S., Møller, B.L., 2009. Substrate specificity of plant UDP-dependent glycosyltransferases predicted from crystal structures and homology modeling. Phytochemistry, doi:10.1016/j.phytochem.2008.12.009.

Paquette, S., Møller, B.L., Bak, S., 2003. On the origin of family 1 plant glycosyltransferases. Phytochemistry 62, 399–413.

Richman, A., Swanson, A., Humphrey, T., Chapman, R., McGarvey, B., Pocs, R., Brandle, J., 2005. Functional genomics uncovers three glucosyltransferases involved in the synthesis of the major sweet glucosides of *Stevia rebaudiana*. Plant J. 41, 56–67.

Ross, J., Li, Y., Lim, E., Bowles, D.J., 2001. Higher plant glycosyltransferases. Genome Biol. 2. Reviews3004.

Sambrook, J., Russell, D.W., 2001. Molecular Cloning, a Laboratory Manual. Cold Spring Harbor Laboratory Press, New York.

Sawada, S., Suzuki, H., Ichimaida, F., Yamaguchi, M.A., Iwashita, T., Fukui, Y., Hemmi, H., Nishino, T., Nakayama, T., 2005. UDP-glucuronic acid: anthocyanin glucuronosyltransferase from red daisy (*Bellis perennis*) flowers. Enzymology and phylogenetics of a novel glucuronosyltransferase involved in flower pigment biosynthesis. J. Biol. Chem. 280, 899–906.

Shao, H., He, X., Achnine, L., Blount, J.W., Dixon, R.A., Wang, X., 2005. Crystal structures of a multifunctional triterpene/flavonoid glycosyltransferase from *Medicago truncatula*. Plant Cell 17, 3141–3154.

Thorsøe, K.S., Bak, S., Olsen, C.E., Imberty, A., Breton, C., Møller, B.L., 2005. Determination of catalytic key amino acids and UDP sugar donor specificity of the cyanohydrin glycosyltransferase UGT85B1 from *Sorghum bicolor*. Molecular modeling substantiated by site-specific mutagenesis and biochemical analyses. Plant Physiol. 139, 664–673.

Thorson, J.S., Barton, W.A., Hoffmeister, D., Albermann, C., Nikolov, D.B., 2004. Structure-based enzyme engineering and its impact on *in vitro* glycorandomization. ChembioChem 5, 16–25.

Unligil, U.M., Rini, J.M., 2000. Glycosyltransferase structure and mechanism. Curr. Opin. Struct. Biol. 10, 510–517.

Vogt, T., Jones, P., 2000. Glycosyltransferases in plant natural product synthesis: characterization of a supergene family. Trends Plant Sci. 5, 380–386.

Weis, M., Lim, E.K., Bruce, N., Bowles, D.J., 2006. Regioselective glucosylation of aromatic compounds: screening of a recombinant glycosyltransferase library to identify biocatalysts. Angew. Chem. Int. Ed. 45, 3534–3538.

Weis, M., Lim, E.K., Bruce, N.C., Bowles, D.J., 2008. Engineering and kinetic characterisation of two glucosyltransferases from *Arabidopsis thaliana*. Biochimie 90, 830–834.

Zhang, Z., Kochhar, S., Grigorov, M., 2003. Exploring the sequence-structure protein landscape in the glycosyltransferase family. Protein Sci. 12, 2291–2302.