Phytochemistry 70 (2009) 325–347

Contents lists available at ScienceDirect

Phytochemistry

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

Review

Substrate specificity of plant UDP-dependent glycosyltransferases predicted from crystal structures and homology modeling

Sarah A. Osmani, Søren Bak, Birger Lindberg Møller *

University of Copenhagen, Department of Plant Biology and Biotechnology, Plant Biochemistry Laboratory, 40 Thorvaldsensvej, DK-1871 Frederiksberg C, Copenhagen, Denmark
VKR Research Centre "Pro-Active Plants", 40 Thorvaldsensvej, DK-1871 Frederiksberg C, Copenhagen, Denmark

ARTICLE INFO

Article history:
Received 17 October 2008
Received in revised form 1 December 2008
Available online 13 February 2009

Keywords:
UGT
Glycosyltransferase
Substrate specificity
Sugar donor specificity
Crystal structure
Homology modeling

ABSTRACT

Plant family 1 UDP-dependent glycosyltransferases (UGTs) catalyze the glycosylation of a plethora of bioactive natural products. In *Arabidopsis thaliana*, 120 UGT encoding genes have been identified. The crystal-based 3D structures of four plant UGTs have recently been published. Despite low sequence conservation, the UGTs show a highly conserved secondary and tertiary structure. The sugar acceptor and sugar donor substrates of UGTs are accommodated in the cleft formed between the N- and C-terminal domains. Several regions of the primary sequence contribute to the formation of the substrate binding pocket including structurally conserved domains as well as loop regions differing both with respect to their amino acid sequence and sequence length. In this review we provide a detailed analysis of the available plant UGT crystal structures to reveal structural features determining substrate specificity. The high 3D structural conservation of the plant UGTs render homology modeling an attractive tool for structure elucidation. The accuracy and utility of UGT structures obtained by homology modeling are discussed and quantitative assessments of model quality are performed by modeling of a plant UGT for which the 3D crystal structure is known. We conclude that homology modeling offers a high degree of accuracy. Shortcomings in homology modeling are also apparent with modeling of loop regions remaining as a particularly difficult task.

Contents

* Corresponding author. Address: University of Copenhagen, Department of Plant Biology and Biotechnology, Plant Biochemistry Laboratory, 40 Thorvaldsensvej, DK-1871 Frederiksberg C, Copenhagen, Denmark. Tel.: +45 353 33352.
E-mail address: blm@life.ku.dk (B.L. Møller).

0031-9422/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.phytochem.2008.12.009

1. Introduction

1.1. Glycosyltransferases (GTs) – a large family of structurally conserved enzymes

Glycosyltransferases (GTs) catalyze the transfer of a sugar residue of an activated sugar donor to an acceptor molecule. The GTs constitute a large family of enzymes. Classification of the GTs into subfamilies has been performed according to the degree of primary sequence identity (Campbell et al., 1997; Coutinho et al., 2003). Members of these families can be found at the CAZY database (http://www.cazy.org/) that at the time of writing comprises 91 different GT families. The primary structure of GTs is poorly conserved (Campbell et al., 1997; Hu and Walker, 2002). Despite low sequence conservation, the GTs crystallized so far show a highly conserved secondary and tertiary structure (Hu and Walker, 2002; Unligil and Rini, 2000; Zhang et al., 2003), and they all adopt distinct folds designated the GT-A and GT-B fold with a third GT-A like fold emerging (Breton et al., 2006). Crystal structure data, as well as sequence analyses argue that the fold structure is conserved among members of the same CAZY family (Bourne and Henrissat, 2001; Breton et al., 2006; Campbell et al., 1997; Coutinho et al., 2003). A second conserved feature of the GTs is the mechanism of the catalytic reaction that is defined as either inverting or retaining. This describes, whether the stereochemistry of the carbon atom of the activated sugar donor engaged in the formation of the new glucosidic bond, is inverted or retained (Coutinho et al., 2003). The catalytic mechanism is generally conserved within a CAZY family, but cannot always be predicted with reliability from sequence comparison alone (Breton et al., 2006). Based on these observations, GTs have been sub-grouped into four clans as either inverting GT-A fold (clan I) or GT-B fold (clan II) or retaining GT-A fold (clan III) or GT-B fold (clan IV) (29). Each CAZY family can be assigned to one of these four clans (Coutinho et al., 2003).

1.2. Family 1 GTs contain the plant UDP-dependent glycosyltransferases (UGTs)

Family 1 GTs are found in all phylae (Campbell et al., 1997). They are inverting GTs and adopt the GT-B fold (Fig. 1A and B) placing this family in clan II. The plant family 1 GTs glycosylates a large array of different small molecules. These include terpenoids, alkaloids, cyanogenic glucosides and glucosinolates as well as flavonoids, isoflavonoids and other phenylpropanoids. Many family 1 GTs utilize an UDP activated sugar as donor in the glycosylation reaction, and most of these belong to a group of GTs referred to as the UGTs (UDP-dependent glycosyltransferases) (Lim and Bowles, 2004; Mackenzie et al., 1997). The UGTs share a highly conserved motif in the C-terminal part denoted the UGT defining sequence. This conserved motif is involved in binding the sugar donor (Mackenzie et al., 1997). A UGT nomenclature based on primary sequence identity of these enzymes has been developed, and positions UGTs with >40% amino acid identity within the same UGT family (Mackenzie et al., 1997). The plant UGTs are assigned to the UGT families 71–100 (Mackenzie et al., 1997), and are further characterized by sharing a highly conserved motif denoted the PSPG motif (Plant Secondary Product Glycosyltransferase motif) (Fig. 2B), that comprises the Prosite UGT defining sequence (Hughes and Hughes, 1994; Paquette et al., 2003). The PSPG containing plant UGTs are soluble enzymes in contrast to the mammalian UGTs that are membrane anchored.

1.3. The importance of plant UGTs

Plants synthesize several hundred thousand bioactive, low molecular mass natural products (secondary metabolites), many of which are glycosylated at specific positions (Jones and Vogt, 2001). Glycosylation alters the stability, solubility, reactivity and biological activity of the aglycons. In grape alone, more than 200

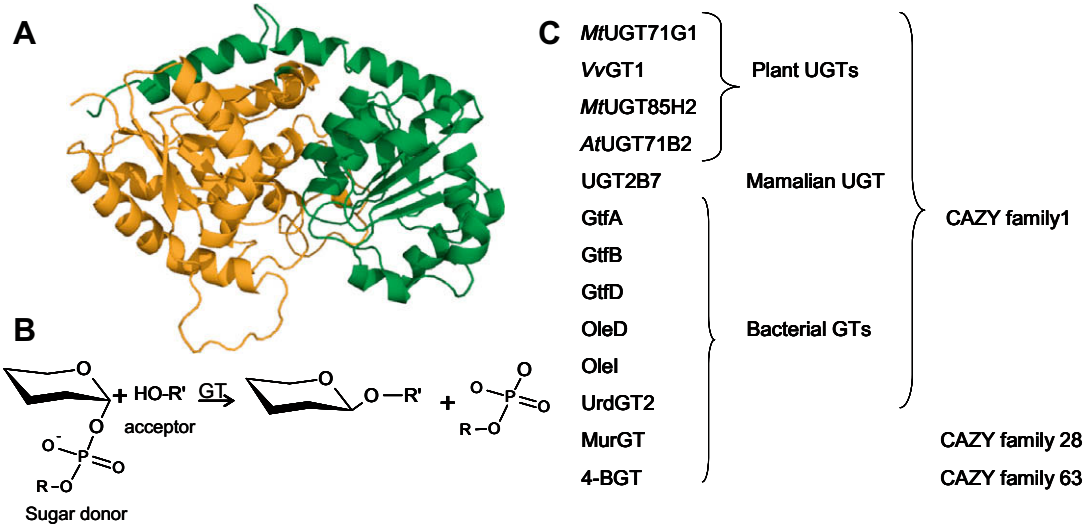


Fig. 1. The characteristics of a glycosyltransferase belonging to clan II. Panel A: Ribbon diagram of a GT adopting the GT-B fold. The N-terminal domain is shown in orange, the C-terminal domain in green. Panel B: The reaction catalyzed by inverting GTs. Panel C: List of inverting GT-B fold GTs with available crystal structures. These will be referred to in the analysis of the plant UGT crystal structures presented in this review. All figures including representations of 3D structures have been made using PYMOL (DeLano, 2002). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

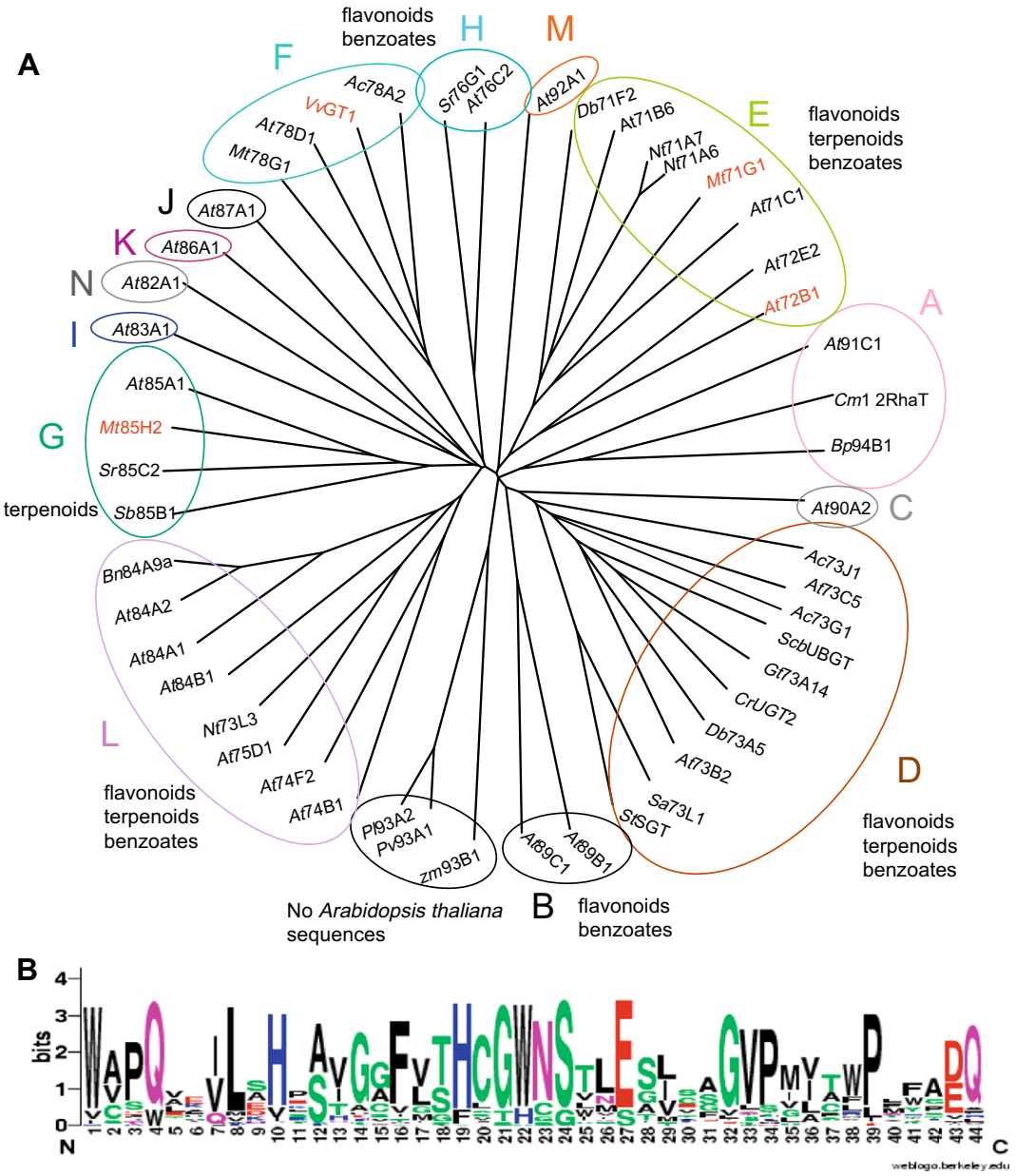


Fig. 2. Panel A: Phylogenetic tree of selected plant family 1 UGTs. Fourteen phylogenetic groups A–N identified in *Arabidopsis thaliana* plus one additional group with no members in *Arabidopsis thaliana* (Bowles et al., 2005; Ross et al., 2001). *At*(*Arabidopsis thaliana*), *Mt*(*Medicago trunkatula*); *Ac*(*Allium cepa*); *Db*(*Dorotheanthus bellidiformis*); *Gt*(*Gentiana triflora*); *Bn*(*Brassica napus*); *Bp*(*Bellis perennis*); *Cm*(*Citrus maxima*); *Cr*(*Catharenthus roseus*); *Nt*(*Nicotiana tabacum*); *pl*(*phaseolus lunatus*); *Pv*(*phaseolus vulgaris*); *Zm*(*zea mays*), *Vv*(*Vitis vinifera*); *St*(*Solanum tuberosum*), *Sr*(*Stevia rebaudiana*), *Sb*(*Sorghum bicolor*). Names of UGTs with solved crystal structures are written in red. Substrate classes (flavonoids, terpenoids and benzoates) glycosylated by at least one *Arabidopsis thaliana* UGT within the different phylogenetic groups are shown (Caputi et al., 2008; Lim et al., 2002, 2004). Panel B: WebLogo of the plant specific PSPG motif constructed from the sequences shown in panel A. Letter size is proportional to the degree of amino acid conservation. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

different glucosides have been identified (Sefton et al., 1994). High plasticity with respect to synthesis of these products enables plants to respond to environmental challenges. In *Arabidopsis thaliana*, 120 UGT encoding genes have been identified (Paquette et al., 2003). Phylogenetic studies have divided these into three clades, two minor clades containing the sterol and lipid GTs and a major monophyletic clade of UGTs that contains the highly conserved PSPG motif (Fig. 2B). The plant UGTs containing the PSPG motif are involved in the synthesis of bioactive natural products, regulation of plant hormone and cell homeostasis as well as in detoxification of xenobiotics (Jones and Vogt, 2001; Lim and Bowles, 2004; Paquette et al., 2003; Vogt and Jones, 2000). It remains an open question how the specificity of these enzymes is controlled. Considering the vast number of glycosides found in plants, high specificity would indicate the requirement for many more UGTs than actually identified, while low specificity would pose problems with respect to undesired side reactions like inactivation of plant hormones (Lim and Bowles, 2004). *In planta* regulation of UGT activity is proposed to involve transcriptional control (Tohge et al., 2005), and incorporation of the UGTs as part of metabolons (Jorgensen et al., 2005; Nielsen et al., 2008), both mechanisms that would help to augment the specificity of the available set of UGTs. Knowledge of the *in vivo* substrate specificity of plant UGTs is important for the detailed understanding of plant plasticity with

respect to natural product synthesis and adaptation to abiotic and biotic environmental stress factors (Lim and Bowles, 2004; Vogt and Jones, 2000).

1.4. Drug development using UGTs

Many bioactive natural products possess pharmaceutically interesting properties in humans (Rates, 2001; Thorson et al., 2001), and 50% of the prescription products sold in Europe and the US are either natural products or natural product derivatives (Cordel, 2002). Position specific glycosylation of these biologically active compounds, serves to modulate their pharmaceutical activity (Kren, 2001; Mijatovic et al., 2007; Weymouth-Wilson, 1997) and/or may serve to alter and optimize their potential use as drug compounds (Bonina et al., 2003; Egleton et al., 2000). Specific glycosylation of complex biomolecules harboring multiple functional groups is difficult to achieve using the technologies of organic chemistry. Enzymatic glycosylation at a desired position can be conferred by proper selection and molecular optimization of UGTs showing the desired regiospecificity. In addition to the use of glycosides in drug development, the health beneficial effects of many fruits and vegetables forming part of the human diet, is thought to reflect their content of bioactive glycosides (Cheynier, 2005; Gelеijnse and Hollman, 2008; Harborne and Williams, 2000; Hollman and Katan, 1997). Accordingly, a detailed knowledge of the *in vivo* and *in vitro* activity and specificity of UGTs would offer the opportunity to design improved UGTs, with desired properties for the production of bioactive molecules for pharmaceutical purposes and to raise the level of health promoting constituents in fruits and vegetables (Bowles et al., 2005; Butelli et al., 2008; Thorson et al., 2004).

1.5. Plant UGTs glycosylate highly divergent acceptors using several different sugars

The enormous variety of glycosides found in plants demonstrates the ability of plant UGTs to glycosylate a large array of highly diverse compounds *in vivo* (Jones and Vogt, 2001). As for the *in vitro* activity of specific UGTs, these show large differences in their individual range of acceptors. Some UGTs are highly specific and catalyze glycosylation of only one or a few acceptors (Fukuchi-Mizutani et al., 2003; Sawada et al., 2005) or acceptor sites (Vogt, 2002; Yonekura-Sakakibara et al., 2007), whereas other UGTs glycosylate a broad range of acceptors *in vitro* (Caputi et al., 2008; Hansen et al., 2003; Hefner and Stockigt, 2003; Kramer et al., 2003; Modolo et al., 2007; Vogt et al., 1997; Vogt and Jones, 2000). An example of a UGT with strict acceptor specificity is the *Gentiana triflora* UGT73A14. This UGT glucosylates delphinidin 3,5-diglucoside at the 3′ position, but has very low activity with delphinidin or delphinidin 3-glucoside, and shows no activity towards other anthocyanins such as the 3,5-diglucosides of pelargonidin or cyanidin, that differ from the delphinidin 3,5-diglucoside only by the number of hydroxyl groups on the B-ring (Fukuchi-Mizutani et al., 2003). In contrast, UGTs such as the *Allium cepa* UGT73G1 (Kramer et al., 2003) and the *Medicago truncatula* UGT78G1 (Modolo et al., 2007) glycosylate a range of different flavonoid aglycones and at several different positions.

Substrate specificity of plant UGTs involves recognition of a sugar acceptor as well as a UDP-sugar donor. The most commonly used sugar is UDP-glucose. Other sugars recognized by plant UGTs include UDP-galactose, UDP-xylose, UDP-mannose and UDP-glucuronic acid. UDP-glucuronic acid is the most common sugar donor used by the mammalian UGTs. In general, the UGTs show very high specificity for the sugar donor (Hefner et al., 2002; Miller et al., 1999; Modolo et al., 2007; Sawada et al., 2005; Yonekura-Sakakibara et al., 2007), but some plant UGTs do show activity with several UDP-sugars, most often with a large preference for one specific UDP-sugar (Jones et al., 2003; Kohara et al., 2007; Modolo et al., 2007; Offen et al., 2006). An example is the *Vitis vinifera* UGT *Vv*GT1 that shows activity with a range of different UDP-sugars, but all with an efficiency lower than 8% compared to that observed with UDP-glucose (Offen et al., 2006). A study of several *M. truncatula* UGTs demonstrates the very high sugar donor specificity observed for most UGTs, as seven of eight tested UGTs showed strict specificity for UDP-glucose while one, *Mt*UGT78G1 exhibited an additional activity towards UDP-galactose. The complexity of sugar donor recognition was further highlighted by the observation that *Mt*UGT78G1 could only use UDP-galactose with a subset of the substrates glucosylated with UDP-glucose (Modolo et al., 2007).

1.6. Substrate specificity cannot be predicted solely based on high amino acid sequence identity

The relationship between the degree of amino acid sequence identity and substrate specificity of the GTs is highly complex and this also applies to the plant UGTs. Highly diverging UGTs are often observed to share the same substrates. This may be illustrated by the two *Dorotheanthus bellidiformis* UGTs UGT73A5 and UGT71F2. These two UGTs show ~20% amino acid sequence identity (Vogt, 2002) but glucosylate the same set of acceptor molecules (Vogt et al., 1997, 1999). The two *A. cepa* UGTs designated *Ac*UGT73G1 and *Ac*UGT73J1 provide an example of the contrary. These two UGTs belong to the same UGT family but the *Ac*UGT73G1 enzyme exhibits very broad acceptor specificity whereas *Ac*UGT71J1 has a very narrow specificity (Kramer et al., 2003).

Phylogenetic studies of the UGTs in *A. thaliana* have defined 14 phylogenetic groups designated A–N (Li et al., 2001; Ross et al., 2001) (Fig. 2A). The inclusion of UGT sequences from other plant species have showed most of these to fall within the defined 14 groups (Bowles et al., 2005; Gachon et al., 2005; Vogt et al., 1997), while an additional group has also been identified (Bowles et al., 2005; Hou et al., 2004) (Fig. 2A). Elucidation of the substrate specificity of the UGTs within these groups has been initiated by screening of the *A. thaliana* UGTs against several acceptors within three major classes; terpenoids (Caputi et al., 2008), benzoates (Lim et al., 2002) and flavonoids (Lim et al., 2004). UGTs within seven of the 14 phylogenetic groups were found to possess activity towards the tested substrates. Among these seven groups, three groups contained UGTs with activity towards all three acceptor classes and three groups had UGTs with activity towards two different acceptor classes (Fig. 2A). This demonstrates that *A. thaliana* UGTs belonging to the same phylogenetically defined group are able to glycosylate acceptors belonging to entirely different classes of compounds, and illustrates the difficulties encountered in attempts to annotate substrate specificity based solely on phylogenetic studies.

A major progress in the understanding of the substrate specificity of UGTs was the concept that UGTs recognize their substrates in a regioselective or regiospecific manner as proposed by Jones et al. (1999) and Vogt and Jones (2000). Regioselectivity can explain why some UGTs glycosylate a broad range of apparently structurally divergent acceptor molecules, which however from a regiospecific point of view harbors structurally similar acceptor sites. A large amount of experimental evidence supports the validity of this concept (Caputi et al., 2008; Jones et al., 1999, 2003; Lim et al., 2002, 2003b, 2004; Modolo et al., 2007; Vogt and Jones, 2000). As an example, the *G. triflora* UGT78B1 has been shown to specifically glycosylate several different anthocyanidins that all present a free hydroxyl group at the 3C position at the C-ring, but differ with respect to the substitution pattern on the B-ring (Tanaka et al., 1996). At the same time it is also evident that presence of a specific accep-

tor site for glycosylation is not enough to secure enzymatic activity and the structure of the entire acceptor molecule is also important. This is observed for the *Clitoria ternatea* UGT UA3'5'GT (39% identity to *At*UGT family 78) that shows regiospecific glycosylation of anthocyanins on the B-ring, while activity is still dependant on the sugar substitution pattern at the A- and C-rings of the anthocyanin (Kogawa et al., 2007).

Phylogenetic analysis of the *A. thaliana* UGTs (Lim et al., 2003a) show that regiospecificity does to a large degree correlate with phylogeny, while some exceptions have to be explained by regioswitching events during evolution (Bowles et al., 2005; Lim et al., 2003a). Accordingly, the ability of distantly related UGTs to glycosylate common acceptors may be explained by their regioselectivity. This is the case for the two distantly related *D. bellidiformis* UGTs UGT73A5 and UGT71F2 mentioned above, that show distinct regiospecificity for their common acceptor betanidin (Vogt et al., 1997). In some cases though, phylogenetically closely related UGTs also show distinct regiospecific differences towards a common acceptor. This is the case for the two *A. thaliana* UGTs *At*UGT74F1 and *At*UGT74F2 that share ~82% amino acid sequence identity and glucosylate the phenolic hydroxyl group of 2-hydroxy benzoic acid. In addition, *At*UGT74F2 is able to catalyze the formation of a glucose ester with this acceptor while *At*UGT74F1 is not (Lim et al., 2002; Lim, 2005). A study of several *M. truncatula* UGTs has also highlighted the difficulties that may be encountered upon assigning substrate specificity based on phylogeny. Biochemical and phylogenetic studies of *Mt*UGT78G1 and *Mt*UGT85H2 showed that substrate specificity could not be predicted by their clustering with biochemically characterized UGTs belonging to the same families (Modolo et al., 2007).

Deduction of substrate specificity not only involves specificity towards the sugar acceptor but also the UDP-sugar donor. Different sugars may be used by UGTs within the same phylogenetically defined families and subfamilies. In cluster F, the UGT78D subfamily UGTs *At*78D2 and *At*78D1 prefer UDP-glucose (Lim et al., 2004) and UDP-rhamnose (Jones et al., 2003), respectively, while the *Aralia cordata* UGT78A2 (Kubo et al., 2004) also belonging to group F (Lim, 2005) uses UDP-galactose.

The prediction of substrate specificity based on phylogenetic analyses will certainly improve as more biochemical data on the substrate preferences of UGTs within each subgroup becomes available (Lim et al., 2003a; Vogt and Jones, 2000). However, for the time being it is evident that substrate specificity cannot be ascribed based on primary sequence analysis alone (Modolo et al., 2007).

Bioinformatics tools for accurate prediction of sequence features important for defining the substrate specificity of a particular UGT have not yet been developed. Progress towards this goal is being made as demonstrated by the recent development of a program to predict acceptors for bacterial GTs that glycosylate antibiotics (Kamra et al., 2005). The program uses a knowledge-based approach and relies on available biochemical and crystal structure data. This means that high amino acid sequence identity between the query sequence and the template, as well as substantial amounts of biochemical data, are essential elements to achieve reliable predictions.

1.6.1. 3D structural data provide a valuable tool for understanding substrate specificity

Biochemical characterization of the substrate specificity of the UGTs is a major task as illustrated by the broad specificity of many UGTs requiring a wide range of substrates to be tested to gain a thorough understanding of the specificity of individual UGTs. When the substrate profile for a UGT has been determined, mutational analyses may provide valuable information regarding the role of single amino acid residues with respect to sugar acceptor and donor specificity and catalytic efficiency. For GTs adopting the GT-A fold, three such studies have identified single amino acid residues determining sugar donor specificity of specific GTs (Marcus et al., 2003; Ouzzine et al., 2002; Ramakrishnan et al., 2005; Ramakrishnan and Qasba, 2002). For the plant UGTs, a mutational study of the *A. cordata* UGT78A2 identified a single amino acid residue responsible for determining UDP-galactose versus dual UDP-galactose and UDP-glucose specificity (Kubo et al., 2004). The inverse mutation in the UDP-glucose specific *Scutellaria baicalensis* UGT UBGT (UGT73 family) did not confer activity with UDP-galactose (Kubo et al., 2004). Mutational studies of *Bellis perennis* UGT94B1 have similarly identified a single amino acid responsible for UGT94B1 activity with UDP-glucuronic acid (Osmani et al., 2008). For more targeted and thus faster biochemical characterization, detailed information of substrate interactions can be obtained from studying protein 3D structures. Such structures can be provided by crystallography. The number of GTs for which the X-ray crystal structure has been solved is limited and reflects the difficulties in obtaining well diffracting crystals and in solving the structures (Breton et al., 2006). In the lack of experimental 3D structures, comparative modeling of proteins using proteins with solved X-ray crystal structures as templates, proposes an alternative approach (Imberty et al., 2006; Kopp and Schwede, 2004). The high conservation of secondary and tertiary structure among the UGTs proposes comparative modeling in the form of homology modeling as an attractive tool for studying their substrate specificity as well as other structure derived properties.

2. 3D structure and substrate specificity of plant UGTs

2.1. A limited number of GT crystal structures are available

The 3D glycosyltransferase database (http://www.cermav.cnrs.fr/glyco3d/) holds crystal structures of 53 different GTs representing a total of 26 different CAZY families. Twenty-six of the GTs with solved crystal structures adopt the GT-B fold and 10 of these belong to the family 1 GTs (listed in Fig. 1). The crystal structure of at least one additional family 1 GT (*At*UGT72B1) may be found at the RCSB protein data bank (http://www.rcsb.org/pdb). Valuable information on enzyme-substrate interactions has been obtained from crystal structures of GTs in complex with different acceptor- or donor-substrates or analogs.

The GT-B fold crystals reveal a two domain structure each domain adopting a α/β/α fold (often referred to as a Rossmann fold) and folded by the N- and C-terminal residues, respectively (Fig. 3B). The C-terminal domain mainly interacts with the sugar donor, while the N-terminal domain mainly interacts with the acceptor.

The crystal structure of a plant UGT has only been resolved recently, the first being *Mt*UGT71G1 (Shao et al., 2005). An additional three plant UGTs have subsequently been crystallized: *Vv*GT1 (Offen et al., 2006), *Mt*UGT85H2 (Li et al., 2007) and *At*UGT72B1 (Brazier-Hicks et al., 2007). The available plant UGT crystal structures are listed in Table 1.

2.1.1. Crystal structures of four plant UGTs reveal their highly conserved 3D structure

From the UGT numbering, three of the crystallized plant UGTs *Mt*UGT71G1, *Mt*UGT85H2 and *At*UGT72B1 are easily identified as belonging to three different UGT families 71, 85 and 72, respectively. The *Vv*GT1 sequence has not been named according to the UGT system yet (Mackenzie et al., 1997). A blast search of the *Vv*GT1 protein sequence against *A. thaliana* UGTs show 55% identity to *At*UGT78D2 placing *Vv*GT1 in the UGT78 family. The four

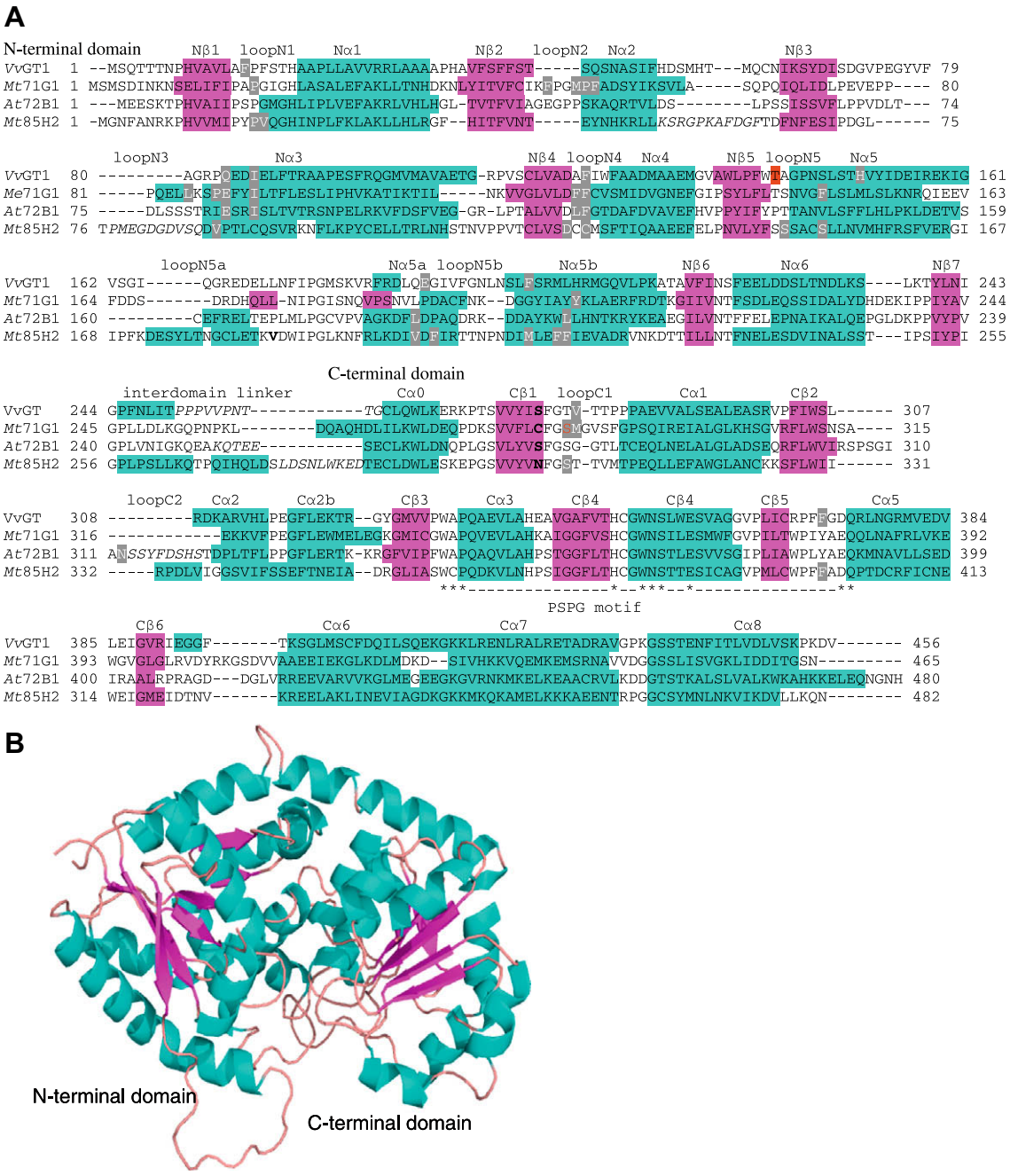


Fig. 3. Panel A: Structural alignment of the four crystallized plant UGTs, α-helices in cyan and β-strands in pink. The alignment was performed by overlay of the 3D structures. The PSPG motif is underlined and 10 conserved sugar donor interacting residues of the PSPG motif are marked with an asterix. Residues reported to form part of the acceptor pocket are highlighted in grey whereas residues outside the PSPG motif reported to interact with the UDP-sugar donor are highlighted in red. Regions in italics indicate regions not clearly defined in the crystal structures. Panel B: Ribbon diagram of the *Vv*GT1 crystal structure showing the 3D folding of elements of secondary structure with α-helices shown in cyan, β-strands in pink and connecting loops in pale pink. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

crystallized plant UGTs are thus from four different UGT families and share 20–35% amino acid identity. In spite of the low amino acid sequence identity, their secondary and tertiary structures are highly conserved (Fig. 3A). Overlay of the coordinates of the crystal structure (Cα's) of *At*UGT72B1 with *Mt*UGT71G1 (402 amino acids), *Vv*GT1 (398 amino acids) and *Mt*UGT85H2 (394 amino acids) gives RMSD values of 1.7 Å, 1.6 Å and 2.0 Å, respectively (Brazier-Hicks et al., 2007).

As expected for family 1 GTs, the four crystallized plant UGTs all adopt the GT-B fold. The 3D-fold and conservation of elements of secondary structure is shown in Fig. 3A. The N-terminal domain folds into a seven-stranded parallel β-sheet surrounded by eight α-helices (numbered Nα1, Nα2, Nα3, Nα4, Nα5, Nα5a, Nα5b and Nα6 Fig. 3), and the C-terminal domain into a six stranded parallel β-sheet surrounded by eight α-helices (numbered Cα0, Cα1, Cα2a, Cα2b, Cα3, Cα4, Cα5, Cα6, Cα7 Fig. 3) and a ninth (Cα8) that folds back over the N-terminal domain. The conserved elements of secondary structure are linked by less conserved loop regions. The loops have been numbered according to the preceding β-strand (except for loops N5a and N5b that follow α-helical structure see Fig. 3A). This numbering is used throughout this review and is shown in Fig. 3A.

Table 1
List of available plant UGT crystal structures, the structures can be found at http://www.rcsb.org/pdb/home/home.do.

Plant UGT	Crystal name	In complex with	Reference
*Vv*GT1	2CIX	UDP	Offen et al. (2006)
	2CIZ	UDP-2-fluoro-glucose, kaempferol	Offen et al. (2006)
	2C9Z	UDP, quercetin	Offen et al. (2006)
*Mt*UGT71G1	2ACV (molecules A and B)	UDP	Shao et al. (2005)
	2ACW (molecules A and B)	UDP-glucose	Shao et al. (2005)
*Mt*UGT85H2	2PQ6	None	Li et al. (2007)
*At*UGT72B1	2VCE	UDP-glucose	Brazier-Hicks et al. (2007)
	2VCH	UDP-Tris buffer	Brazier-Hicks et al. (2007)
	2VG8	UDP-2-fluoro glucose, TCP	Brazier-Hicks et al. (2007)

The most diverging of the loops with respect to both length and sequence are the loops N3, N5 and N5a all positioned in the N-terminal domain, as well as loop 7 that constitutes the linker region connecting the two domains (interdomain linker). In some cases these regions contain elements of unconserved secondary structure. A short α-helix is observed in the N5 loop of *Mt*UGT85H2 and *At*UGT72B1 as well as in the interdomain linker region of *Mt*UGT85H2 and *Vv*GT1, while *Mt*UGT71G1 has two short stretches of β-sheet in the N5 loop (Fig. 3A).

2.2. Substrate accommodation involves structural changes

The crystal structures of the four plant UGTs show that the two domains pack very tightly resulting in the formation of a deep narrow cleft constituting the substrate pocket (Fig. 4). The narrow access to the binding pocket suggests that substrate accommodation occurs following cleft opening. Several observations of the plant UGT crystals support that domain movements are mediated by flexibility of the linker region. The unit crystal of *Mt*UGT71G1 contains two independent UGT molecules, and small differences between the two, especially in the cleft region, suggest domain movement (Shao et al., 2005). This is further supported by the observation that in one of the unit crystal molecules, the linker region connecting the N- and C-terminal domains is poorly defined. This is also the case in both the *Vv*GT1 (Offen et al., 2006) and *Mt*UGT85H2 (Li et al., 2007) crystals, where disordered structure of this region suggests flexibility. Domain movements governed by more or less flexible linker regions are observed for many multi domain proteins (Wriggers et al., 2005). Changes in the relative position of the N- and C-terminal domain upon substrate binding is also proposed for other crystallized GTs (Gibson et al., 2004; Ha et al., 2000; Mulichak et al., 2003; Qasba et al., 2005; Unligil and Rini, 2000). Domain movement is proposed to occur without changing the 3D structure of the structurally conserved regions of the individual domains (Wriggers et al., 2005), as observed for the bacterial GT-B fold proteins MurG (Ha et al., 2000) and GtfA (Mulichak et al., 2003), where cleft opening proceeds by a small rigid movement in the linker region displacing the two domains relative to each other.

In addition to flexibility of the linker region, structural flexibility in other loop regions associated with substrate binding appears to be a general feature of GTs (Qasba et al., 2005; Unligil and Rini, 2000), and is suggested from crystal structure studies of several GT-B fold proteins (Bolam et al., 2007; Gibson et al., 2004; Ha et al., 2000; Li et al., 2007; Mittler et al., 2007; Morera et al., 1999; Mulichak et al., 2003, 2004). Another observation supporting the occurrence of small structural changes in the plant UGTs upon substrate binding is the differences observed in the positioning of a conserved tryptophan (W) residue located as the first residue of the PSPG motif. In the *Mt*UGT85H2 crystal formed in the absence of any substrate, the indol ring system of W360 is positioned differently from the position in three other crystals, where stabilization of the uracil ring of the UDP-sugar donor is mediated by the W360 residue (Fig. 5). Structural changes in the positioning of the W360 residue upon binding of sugar donor analogs are also suggested by comparing to the crystal structures of two other family 1 UGTs Ole1 from *Streptomyces antibioticus* crystallized in complex with UDP (Bolam et al., 2007) and to UGT2b7 from *Homo sapiens* (partly) crystallized in absence of substrates (Miley et al., 2007). In these UGTs, positioning of the conserved tryptophan side chain differs accordingly in the presence or absence of the donor substrate. Structural changes in this nucleotide accommodating region are also observed upon nucleotide-sugar binding in the GT-B fold enzymes GtfA (Mulichak et al., 2003), GtfD (Mulichak et al., 2004) and murG (Hu et al., 2003), and are thought to reflect domain closure (Mulichak et al., 2003).

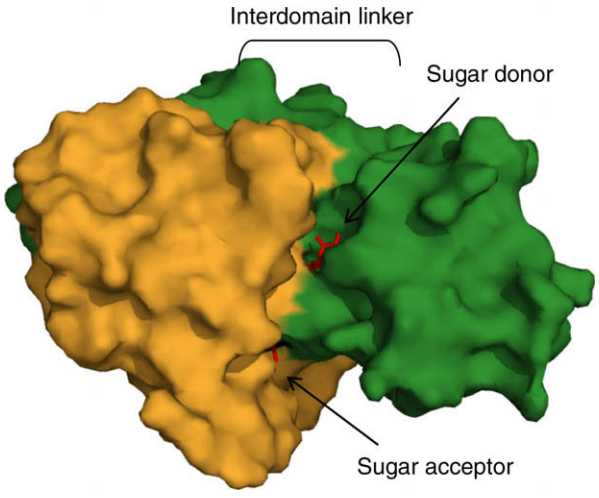


Fig. 4. Surface view of the *Vv*GT1 crystal showing the buried state of the sugar donor and acceptor (red stick) between the N-terminal (orange) and C-terminal (green) domains. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

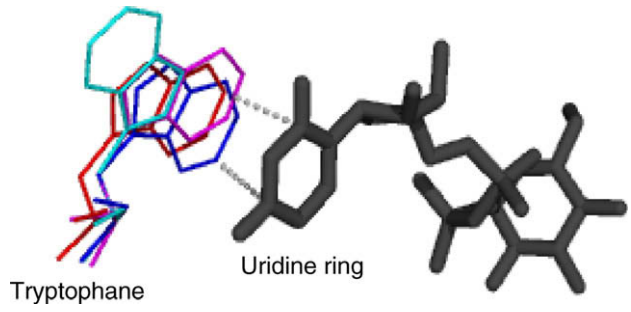


Fig. 5. Overlay of the 3D structure of the four plant crystals showing the relative position of the first PSPG motif residue, a conserved tryptophane (W), and the uridine moiety of the UDP-sugar donor. *Mt*UGT85H2 (W360 cyan), *Vv*GT1 (W332, blue), *Mt*UGT71G1 (W339, red) and *At*UGT72B1 (W351, purple). The shown positioning of the UDP-sugar donor is based on the *Vv*GT1 crystal structure (black). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.3. Substrate specificity is determined by several regions within the primary sequence

2.3.1. The sugar donor pocket

2.3.1.1. Amino acids of the PSPG motif provides the majority of sugar donor interactions. The sugar donor pocket is mainly formed by residues from the C-terminal domain. One side of the pocket is flanked by the PSPG motif (Fig. 6A) with the highly conserved Rossmann fold featuring an α/β/α structure where residues in the N-terminal end of the α-helices Cα3, Cα4 and Cα5 (/loop C5) are positioned pointing towards the UDP-sugar donor. Both the length and residues of these regions are highly conserved between UGTs as are the potential interactions between the UGTs and the UDP-sugar donor. The crystal structures of *Mt*UGT71G1, *Vv*GT1 and *At*UGT72B1 have been solved with the sugar donor (or part of it, see Table 1) and 10 highly conserved residues of the 44 amino acid PSPG motif are observed to directly interact with the UDP-sugar (marked with an asterix in Fig. 3A). The same interactions are predicted in the *Mt*UGT85H2 crystal by docking of UDP-glucose (Li et al., 2007). The additional conserved residues that do not interact directly with the sugar donor may serve to stabilize or facilitate intramolecular interactions. In *Mt*UGT85H2, several interdomain interactions involve PSPG residues (Li et al., 2007).

Of the 10 sugar donor interacting PSPG residues, seven interact with the invariant part of the UDP-sugar donor. The last three residues are the conserved W (22nd amino acid residue of the PSPG motif) (Fig. 2) and the two C-terminal residues of the PSPG motif conserved as D/E/S and Q/E/H//N, respectively. In the four crystallized plant UGTs, these last two residues are aspartic acid (D)/glutamic acid (E) and glutamine (Q), respectively (Fig. 6C). These three residues W, D/E and Q are positioned to form hydrogen bonds to the sugar part of the donor. Whereas the residues interacting with the invariant part of the UDP-sugar were expected to be highly conserved, the residues of the PSPG motif interacting with the variable sugar moiety might have been expected to vary dependent on the type of sugar. This is not the case. Accordingly, additional amino acid residues are involved in defining the specificity towards a particular sugar (Kubo et al., 2004; Offen et al., 2006; Shao et al., 2005). The availability of four plant UGT crystal structures provides a new starting point to assess the importance of the individual residues in the PSPG motif in recognition of the UDP-sugar.

For the four crystallized plant UGTs, the preferred sugar donor is UDP-glucose and the position of the three PSPG residues interacting with the sugar in the four crystals are positioned almost exactly identically. H-bonds between the hydroxyl groups of the glucose residue and the three amino acid residues are observed as follows: 4′OH and the W backbone, 3′OH and 4′OH and the D/E residue, and 2′OH and 3′OH and the Q residue (Fig. 6B). In galactose, the 4′OH group points in the opposite direction from that in glucose. This has been proposed to prevent formation of stabilizing bonds to the D/E residues (Shao et al., 2005). Activity with UDP-galactose is indeed not observed for *Mt*UGT85H2 (Li et al., 2007), while low activity is observed for *Mt*UGT71G1 (1–4% of UDP-glucose) (Shao et al., 2005) and *Vv*GT1 (8% of UDP-glucose) (Offen et al., 2006). However, the importance of this residue for galactose activity is questioned by the fact that the D/E residue is highly conserved in galactosyltransferases (Kubo et al., 2004). Detailed examination of the crystal structures does indicate that the D/E residue is able to from an H-bond with the 4′OH of galactose and that interaction to the W residue is more likely to be abolished. In *A. thaliana* flavonoid UGTs, the D/E residue is conserved as an aspartic acid (D) (Yonekura-Sakakibara et al., 2007) even though these UGTs show specificity for a variety of sugars differing in conformation, thus contradicting that this amino acid residue determines sugar specificity (Yonekura-Sakakibara et al., 2007). Point mutation of the aspartic acid (D)/glutamic acid (E) residue to alanine (A) in

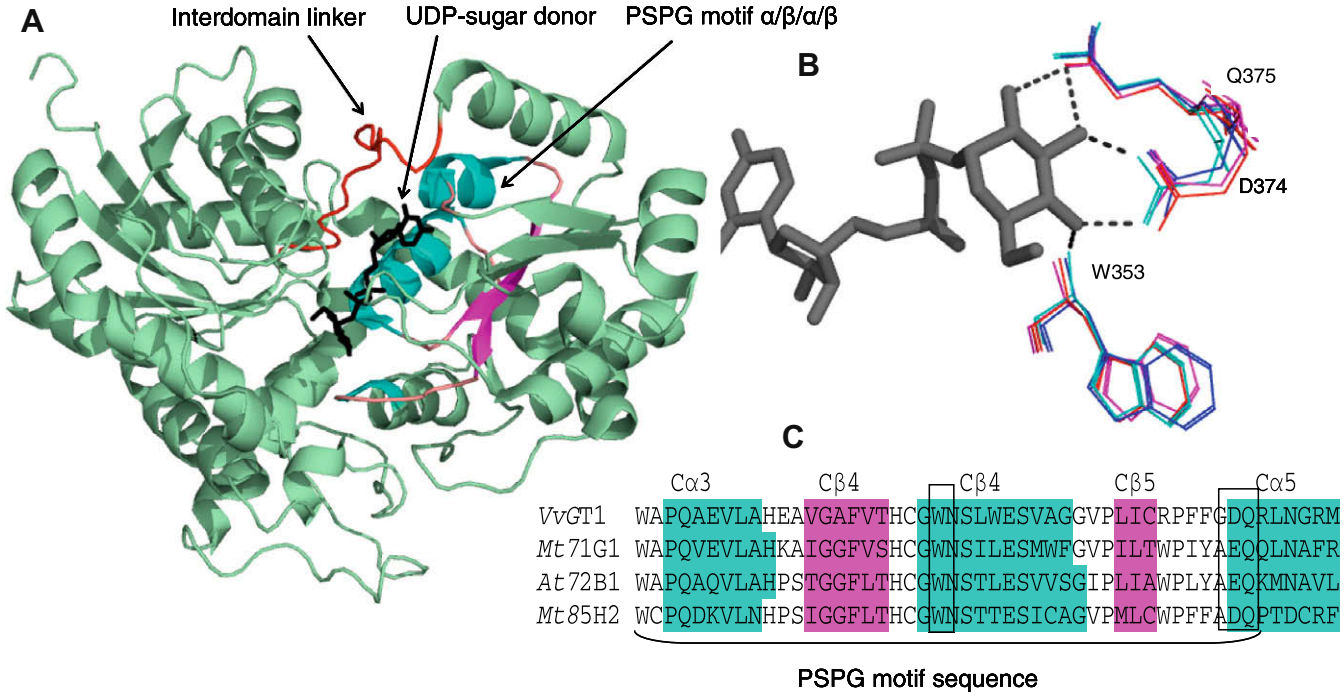


Fig. 6. Position of the UDP-sugar donor. Panel A: Ribbon diagram of the *Mt*UGT71G1 crystal. The UDP-sugar donor (black stick model) is flanked by the PSPG motif on one side (α-helices cyan and β-strands pink) and the interdomain linker (red) on the other side. Panel B: Overlay of conserved sugar interacting residues (W, D/E and Q) from the four plant crystals, showing interactions with the sugar part of the UDP-sugar donor. Color code: *Vv*GT1 blue, *Mt*UGT71G1 red, *Mt*UGT85H2 cyan, *At*UGT72B1 purple, UDP-glucose black. Panel C: Alignment of their PSPG motif sequence with the three sugar interacting residues boxed (W, D/E and Q). α-Helices in cyan and β-strands in pink. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

*Mt*UGT71G1 (Shao et al., 2005) and *Vv*GT1 (Offen et al., 2006) abolished activity demonstrating the overall importance of this residue.

The last residue in the PSPG motif in the four crystallized plant UGTs is glutamine (Q) (Fig. 6, panel C). *Vv*GT1 can use both UDP-glucose and UDP-galactose as sugar donors whereas *Mt*UGT85H2 cannot. The involvement of the Q residue in determining sugar specificity has been studied in the *A. cordata* galactosyltransferase UGT78A2 (Kubo et al., 2004) that carries a histidine (H) residue at this position. The mutation H374Q switched sugar donor preference from UDP-galactose to UDP-glucose. The inverse mutation Q382H in the *S. baicalensis* UGT UBGT (UGT73 family) did not switch specificity from UDP-glucose to UDP-galactose (Kubo et al., 2004). In *Vv*GT1, mutation of the glutamine (Q) into a histidine (H) residue abolished activity with UDP-glucose as well as UDP-galactose (Offen et al., 2006). Rhamnose has a 4′OH configuration similar to that of galactose. In the rhamnosyltransferases *At*UGT78D1 and *At*UGT89C1 of *A. thaliana* the last residue in the PSPG motif is occupied by an aspartic acid (D) and a histidine (H) residue, respectively (Yonekura-Sakakibara et al., 2007). SGT (UGT73 family) from *Solanum tuberosum* possesses a dual glucosyl- and galactosyltransferase activity whereas UGT73L1 from *Solanum aculeatissimum* has only glucosyltansferase activity. Both UGTs harbor a glutamine (Q) residue at the last position of the PSPG motif (Kohara et al., 2007). Domain swap studies between C-terminal loop regions of *Sa*UGT73L1 and *St*SGT have been conducted, but did not identify specific residues enabling activity towards UDP-galactose in *St*SGT (Kohara et al., 2007). As previously stated, the preferred sugar donor of all four crystallized plant UGTs is UDP-glucose. The C-terminal domain (amino acid residues 285–446) of human UGT2B7 that uses UDP-glucuronic acid as sugar donor has also been crystallized (Miley et al., 2007). In this crystal, the position of the D/E and Q residues in the 3D structure (conserved as aspartic acid (D) and glutamic acid (Q) in UGT2B7) is identical to the plant crystals. Other residues involved in determining the sugar specificity of this UGT were not resolved. In conclusion, sugar donor specificity is dependent on the last two residues in the PSPG motif and on interaction with a number of other yet unknown residues.

2.3.1.2. The interdomain linker could be important for sugar donor binding. The interdomain linker region flanks the UDP-sugar donor opposite the PSPG motif (Fig. 6A). Part of this region is distorted in the *Mt*UGT85H2, the *Vv*GT1 and the *At*UGT72B1 crystals indicating flexibility of this region. In the *Mt*UGT71G1 crystal structure,

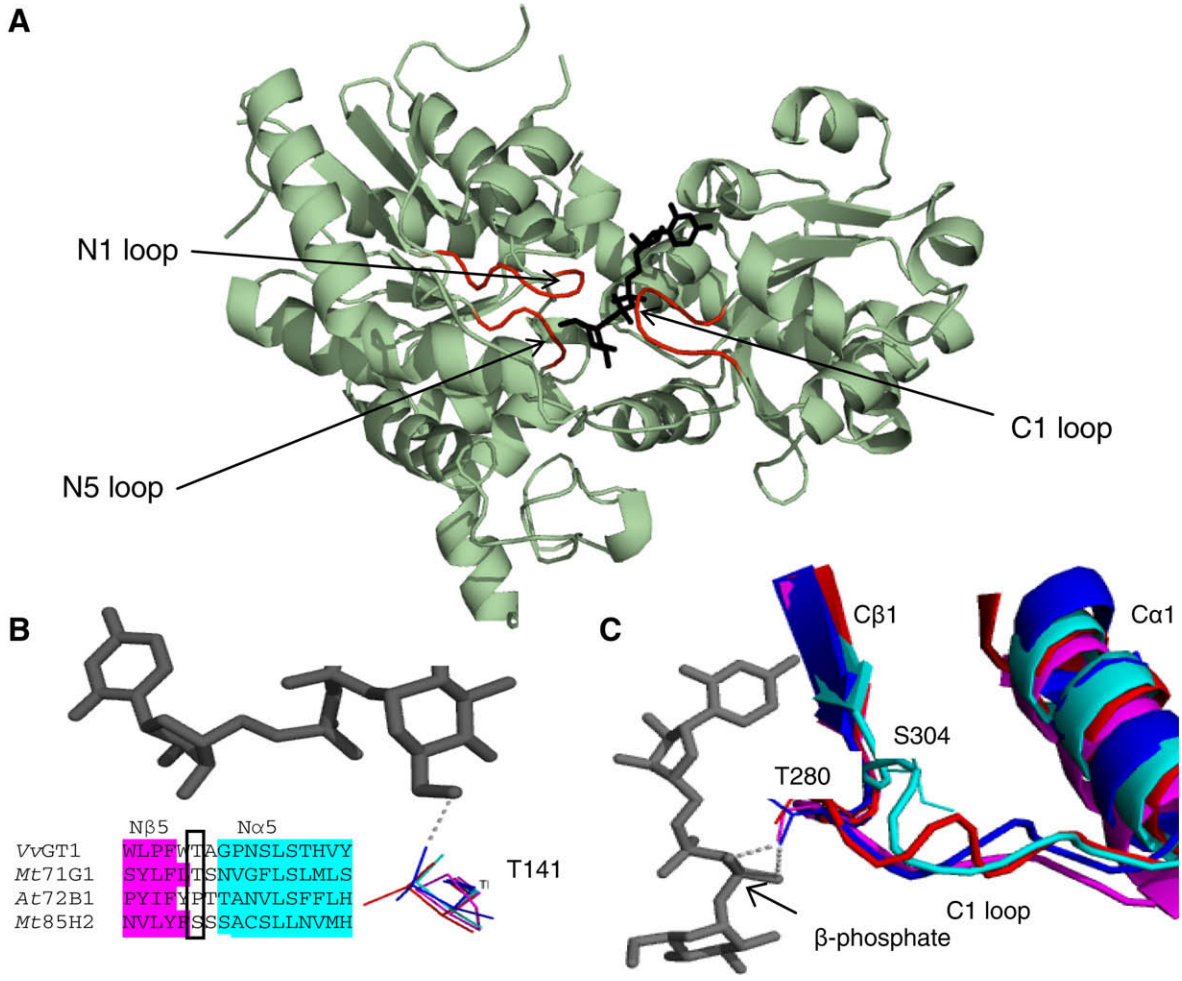


Fig. 7. Regions outside the PSPG motif offering sugar-donor interactions. Panel A: Ribbon diagram of the crystal structure of *Vv*GT1 showing the position of the N5, N1 and C1 loops (red) and of UDP-glucose (black stick model). Panel B: The N5 loop illustrating residue T141 from *Vv*GT1 (blue) that is proposed to interact with the glucose 6′OH of UDP-glucose. The corresponding residues in *Mt*UGT71G1, *Mt*UGT85H2 and *At*UGT72B1 are shown by a structural alignment of the region, and their similar position in the 3D structure by lines *Mt*UGT71G1 (T143 red), *Mt*UGT85H2 (S147 cyan) and *At*UGT72B1 (P139 purple) UDP-glucose from *Vv*GT1 is shown in black. Panel C: The C1 loop with the preceding Cβ1 and the following Cα1, residues T280 from *Vv*GT1 and S304 from *Mt*UGT85H2 are named and shown in stick along with corresponding S285 from *Mt*UGT71G1 and S277 from *At*UGT72B1 showing the potential stabilizing interactions with the β-phosphate of UDP. Color code: blue, *Mt*UGT71G1 red, *Mt*UGT85H2 cyan, *At*UGT72B1 purple, UDP-glucose from *Vv*GT1 in black. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

where the entire linker region can be defined in one of the two unit crystals, the C-terminal part of the linker is positioned close to the uridine moiety of the UDP-sugar donor (Fig. 6A), but not close enough to form any direct interactions. Intramolecular interactions involving linker residues might be of importance, and the linker loop is positioned close to the W339 predicted to interact with the uridine moiety of the UDP-sugar donor, and suggested to undergo conformational change upon sugar donor binding (Fig. 5). In a mutational study of *Mt*UGT85H2, deletion of the residues 274–282 of the linker reduced activity to 6% of wild-type. This demonstrates the importance of this region for activity (Li et al., 2007).

The linker region of plant UGTs is highly variable with respect to length and sequence. In the four crystallized plant UGTs, the linker spans from 15 amino acids in *Mt*UGT71G1 to 28 in *Mt*UGT85H2 (end of Nβ7 to start Cα0 see Fig. 3). This region also differs considerably when comparing to bacterial family 1 GTs, and in the crystallized *Amycolatopsis orientalis* GtfA, GtfB and GtfD it is much shorter spanning only 13 amino acids. The crystal structures of GtfA, GtfB and GtfD show direct H-bond interactions between linker residues and the sugar donor (Mulichak et al., 2003, 2004). This is not observed in the plant crystals. Differences in length and composition of the linker region thus serve to define domain positioning and substrate accommodation.

2.3.1.3. Important interactions with the UDP-sugar donor are provided by residue in the C1 loop. C-terminal residues outside the PSPG motif that interacts with the sugar donor are found in the C1 loop (Fig. 7A and C). The length of the C1 loop is conserved in the plant UGTs apart from an additional residue found in *Mt*UGT71G1. In bacterial and many other GT-B fold GTs, a short sequence motif within this loop is conserved (Ha et al., 2000; Hoffmeister et al., 2001; Hu et al., 2003). This motif harbors a conserved glycine followed by serine/threonine/arginine. In the four plant UGT crystals, this motif is present as glycine followed by serine/threonine. Comparing the 3D position of this loop in the four plant UGT crystals, the position is similar in the crystals of *Mt*UGT71G1 (Shao et al., 2005), *Vv*GT1 (Offen et al., 2006) and *At*UGT72B1 (Brazier-Hicks et al., 2007) co-crystallized with the UDP-part of the sugar donor (Fig. 7C). In these crystals, the conserved serine/threonine is positioned to interact with the sugar donor β-phosphate. In the *Mt*UGT85H2 crystal structure, where UDP is absent, the loop is positioned slightly further away and interaction between the serine and β-phosphate would require loop movement (Fig. 7C). This implies mobility of this loop upon sugar donor binding. Conservation of the interaction with the β-phosphate group is supported by observations on the crystal structures of the *Escherichia coli* GT-B fold GT MurG and the *A. orientalis* family 1 GT GtfA, where the conserved serine residue of the C1 loop also is positioned to interact with the β-phosphate of the UDP-sugar donor and where mobility of the loop accompanies substrate binding (Hu et al., 2003; Mulichak et al., 2003).

Direct involvement of C1 loop residues in UDP-sugar donor specificity has been observed in a mutational study of *Mt*UGT71G1. Mutation of the C1 loop residue M286 into leucine caused increased activity with the alternative sugars UDP-galactose and UDP-glucuronic acid. No explanation for the changed specificity was proposed (He et al., 2006), but this residue is positioned adjacent to the above mentioned serine residue involved in UDP β-phosphate interaction and close to the sugar moiety of the donor.

2.3.1.4. N-terminal domain residues are also involved in determining sugar donor specificity. Two regions of the N-terminal domain are positioned close to the sugar donor, and can propose interactions with the sugar moiety. These are the loops N1 and N5 (Fig. 7A). These loops are conserved in length and the N1 loop terminates with a highly conserved histidine residue (Fig. 3A). This residue has been shown to be crucial for the catalytic activity of UGTs (Brazier-Hicks et al., 2007; Hans et al., 2004; Li et al., 2007; Offen et al., 2006; Shao et al., 2005). Apart from this conserved histidine, the residues of the N1 and N5 loops are not highly conserved. When the residue T141 in *Vv*GT1 positioned in the N5 loop region was mutated (T141A), a sixfold decrease in activity was observed (Offen et al., 2006). From the crystal structure it is apparent that this residue interacts with the 6′OH of the sugar donor (Fig. 7B). This residue is often conserved as threonine or serine, and interaction of this residue with the sugar donor is also predicted for the crystals of *Mt*UGT71G1 and *Mt*UGT85H2, while a proline takes this position in *At*UGT71B2. Importance of N-terminal residues for sugar donor specificity has also been observed in a domain swap study of the UGTs UGT93A2 (ZOG1) from *Phaseolus lunatus* and UGT93A1 (ZOX1) from *Phaseolus vulgaris* (Martin RC, 2000). In this study an N-terminal region encompassing the N5 loop was observed to be decisive for specificity for UDP-glucose, and offers good targets for more detailed studies on sugar donor specificity.

In the *B. perennis* glucuronosyl transferase *Bp*UGT94B1, molecular modeling proposed that R25 positioned in the N1 region interacts with the negatively charged 6′-carboxylate of the sugar donor UDP-glucuronic acid. This arginine residue is not conserved in UGTs but the importance of the residue for sugar donor recognition in UGT94B1 was confirmed by site directed mutagenesis. The point mutant R25S shows a 100-fold decrease in activity with UDP-glucuronic acid, whereas the activity with UDP-glucose increased threefold (Osmani et al., 2008) (Fig. 11).

In conclusion, these studies demonstrate that although the C-terminal provides the majority of the residues interacting with the sugar donor, important substrate interactions involve residues positioned within the N-terminal domain.

2.3.2. The sugar acceptor pocket

2.3.2.1. The acceptor pocket is formed by less conserved regions. - Whereas the sugar donor mainly interacts with residues from the C-terminal domain, the sugar acceptor pocket is almost entirely formed by N-terminal residues. The N-terminal domain is less conserved than the C-terminal domain. In the structural alignment (Fig. 3A), residues described to form part of the acceptor pockets in the four plant UGT crystals are highlighted in grey (Brazier-Hicks et al., 2007; Li et al., 2007; Offen et al., 2006; Shao et al., 2005). The pocket is mainly formed by residues in the less conserved loop regions that connect elements of conserved secondary structure. These are the loops N1, N2, N3 (including the beginning of Nα3), N4 and N5 including the helices Nα5, Nα5a, Nα5b and the loops connecting them (N5a and N5b) (Fig. 8A). Although the acceptor pocket is mainly formed by N-terminal residues, some C-terminal residues contribute to the formation of the pocket. Analysis of the plant UGT crystals locates these residues to the C1 and the C5 loops (Fig. 8A), loops that are also known to interact with the sugar donor regions. The C1 as well as C5 loops are conserved in length and also show conserved backbone structure in the four crystals. The C5 loop constitutes the final part of the PSPG motif.

2.3.2.2. Correct acceptor positioning is a prerequisite for activity. A crucial point for activity is the position of the acceptor –OH, $–NH_2$ or –SH functional groups amenable to glycosylation. For inverting GTs, the accepting functional group needs to be positioned near the 1′C of the sugar-donor glycose and near to the amino acid acting as the general base facilitating deprotonation of the acceptor. In the majority of plant UGTs, this amino acid is a histidine residue (Brazier-Hicks et al., 2007; Li et al., 2007; Offen et al., 2006; Shao et al., 2005). The correct orientation of these

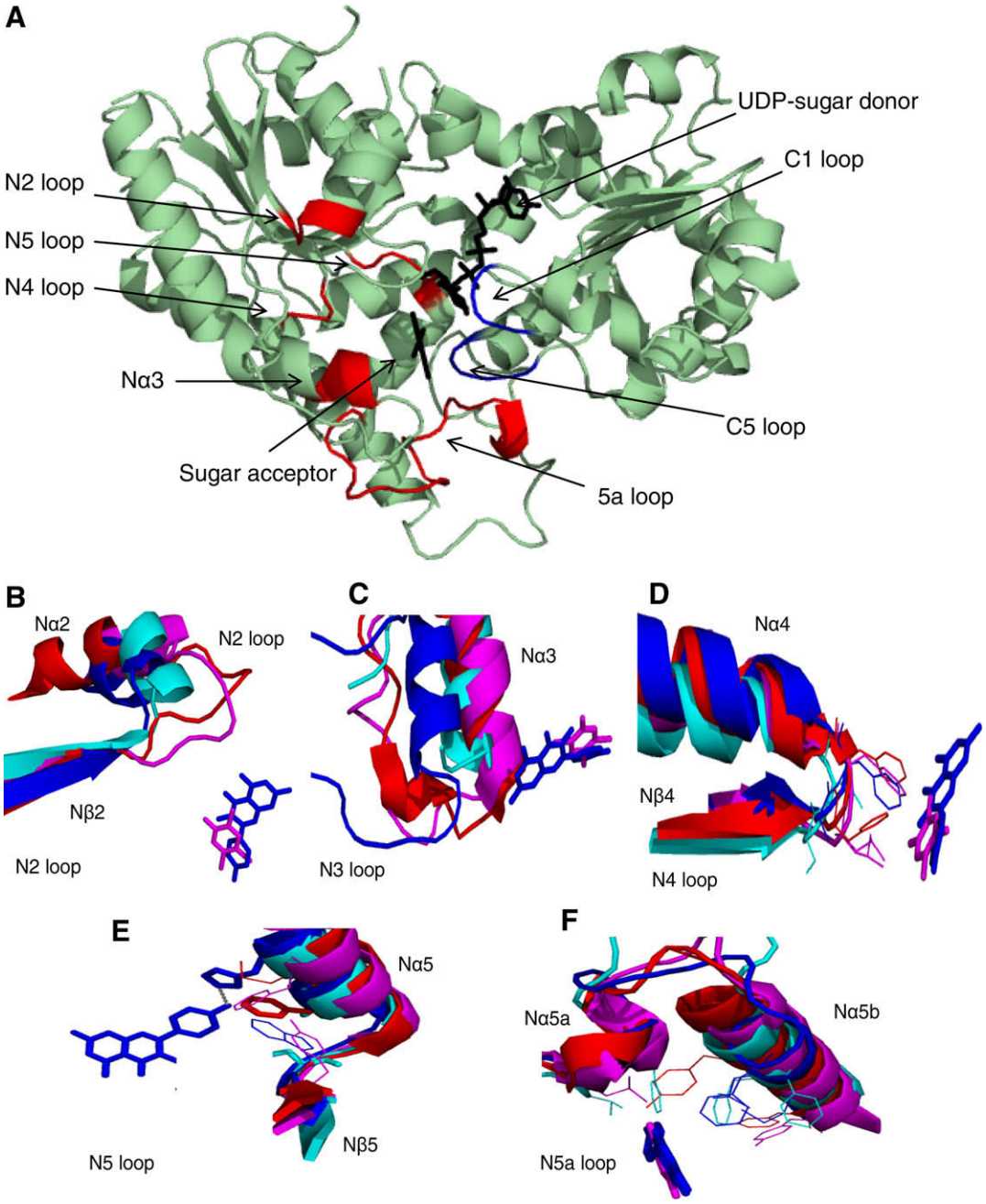


Fig. 8. Regions forming the sugar acceptor pocket. Panel A: Ribbon diagram of the crystal structure of *Vv*GT1. The regions in the N-terminal that form part of the sugar acceptor pocket are shown in red and the C-terminal regions forming part of the sugar acceptor pocket in blue. The UDP-sugar donor and acceptor are shown as black stick models. Panels B–F: Ribbon structure overlays of the four plant crystals showing the N-terminal regions forming the acceptor pocket. Side chains of large amino acid residues pointing into the pocket are shown as stick models. Panel B: N2 loop. Panel C: Nα3. Panel D: N4 loop. Panel E: N5 loop. Panel F: N5a loop. Color code: *Vv*GT1 blue, *Mt*UGT71G1 red, *Mt*UGT85H2 cyan, *At*UGT72B1 purple. The sugar acceptor kaempferol from *Vv*GT1 is shown in a blue stick model and sugar acceptor TCP from *At*UGT72B1 in purple. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

functional groups determines whether an enzymatic reaction can take place. A large number of different acceptor molecules may offer the same functional group that can be correctly positioned into the active site explaining the observed regiospecificity of plant UGTs (Vogt and Jones, 2000). Correct positioning in the active site of the enzyme is the first prerequisite for activity and should be compatible with accommodation of the rest of the acceptor molecule and positioning of the activated sugar donor molecule. Differences in both backbone structure as well as in amino acid side chains that offer stabilizing interactions or provide space restrictions, determine which acceptor molecule(s) can be accommodated by a specific UGT.

2.3.2.3. The backbone structure determines the overall shape of the acceptor pocket. Superimposing the peptide backbone chains of the regions forming the acceptor pockets in the four plant UGT crystal structures show that despite poor amino acid conservation within these regions, the backbone structure of many of them is actually highly conserved. Two regions that diverge with respect to backbone position are regions surrounding the N2 loop and the Nα3

(Fig. 8A). The N2 loop regions of *Mt*UGT71G1 and *At*UGT72B1 are situated close to the acceptor (Fig. 8B). This loop is not found in *Mt*UGT85H2 and *Vv*GT1 where the Nα2 helix follows straight after Nβ2 (Fig. 8B). In *Mt*UGT71G1, docking studies have proposed stabilizing H-bond interactions of residues within this region to the acceptors medicagenic acid and hederagenin, while the smaller acceptor molecule kaempferol is positioned further away from these residues (Shao et al., 2005). Analyses of the *Vv*GT1 and *At*UGT72B1 crystal structures with the acceptors kaempferol and TCP, respectively, show that residues within this region are positioned too far from the acceptor to enable direct interactions (Fig. 8B).

The N3/Nα3 region lines part of the acceptor pocket in all four crystal structures but differs in length and position (Fig. 8C). Nα3 is positioned closer to the acceptor in *Mt*UGT71G1 and *At*UGT72B1 than in *Vv*GT1, while it takes an intermediary position in *Mt*UGT85H2. The N3 loop of *Mt*UGT85H2 is poorly defined in the crystal, but the residues preceding Nα3 are positioned closer to the acceptor site and thereby restricts the size of the pocket as compared to *Vv*GT1.

An examination of the six crystallized bacterial family 1 GTs (listed in Fig. 1C) identifies the N3–Nα3 loop region as one of two hypervariable loops in the N-terminal domain engaged in determining acceptor specificity (Bolam et al., 2007; Mittler et al., 2007; Mulichak et al., 2001, 2003, 2004). The *S. antibioticus* UGTs OleD and OleI show an amino acid sequence identity of 46% and differs in acceptor specificity. The crystal structures of these enzymes show structural variations in the acceptor site to be caused by differences in only two regions, one of them being the loop N3–Nα3 (Bolam et al., 2007). An exchange of the N3–Nα3 loop between the two UGTs did not switch specificity but resulted in decreased activity emphasizing the complex nature of substrate recognition (Bolam et al., 2007). Further mutational studies on OleD provided a triple point mutant with much broader substrate specificity. The most contributing single mutation was located within the N3 loop (Williams et al., 2007). The importance of residues within the N3 loop with respect to substrate specificity has also been shown in studies with the *Streptomyces fradiae* UGTs UrdGT1a and UrdGT1c (Hoffmeister et al., 2001, 2002). These two UGTs share 91% amino acid sequence identity. Domain swaps and introduction of point mutations ascribed differences in acceptor and donor specificity to the region corresponding to loop N3–Nα3 (Hoffmeister et al., 2001, 2002).

The observed differences in the position of the N3–Nα3 loop in plant UGTs and the involvement of this region in determining acceptor specificity in other family 1 GTs show that amino acid residues positioned in this loop are key determinants of acceptor specificity of plant UGTs.

2.3.2.4. The specific features of the acceptor pocket are determined by the amino acid environment. The acceptor pockets in the four plant UGT crystal structures are mainly lined with residues creating a hydrophobic environment. In some cases, amino acid residues proposing direct interactions in the form of H-bond formation with the acceptor are observed. In *Vv*GT1 and *Mt*UGT71G1 different stabilizing interactions are proposed for different acceptors (He et al., 2006; Li et al., 2007; Offen et al., 2006). As previously outlined, many of the regions forming the acceptor pocket possess nearly identical backbone structures. An important factor determining differences in the shape of the acceptor pocket then becomes the size of amino acid side chains pointing into the pocket. Difference in size and properties of these amino acids are observed to be of crucial importance for acceptor specificity (He et al., 2006).

Amino acid residues influencing the size of the acceptor pocket are observed to differ in three of the loop regions forming the acceptor pockets in the four plant UGT crystal structures. The N4 loop of *Mt*UGT71G1 contains two phenylalanine residues. The corresponding two residues in *Vv*GT1, *At*UGT72B1 and *Mt*UGT85H2 are alanine and phenylalanine, leucine and phenylalanine, and two cysteine residues, respectively (Fig. 8D). Large side chains in the N5 loop and Nα5 are histidine and tryptophan (H150 and W140) in *Vv*GT1, phenylalanine (F148) in *Mt*UGT71G1 and phenylalanine and tyrosine (F148 and Y138) in *At*UGT72B1, while none are found in *Mt*UGT85H2 (Fig. 8E). In Nα5a and Nα5b of *Mt*UGT71G1 and *Vv*GT1 one tyrosine residue (Y202 and Y200, respectively) points into the acceptor pocket. In *Mt*UGT85H2 three tyrosine residues (Y202, Y215 and Y216) as well as a methionine (M212) are pointing into the pocket in this region, while in this region of *At*UGT72B1 there are no residues with large side chains (Fig. 8F).

Biochemical studies of the acceptor specificity of *Mt*UGT71G1, *Mt*UGT85H2 and *Vv*GT1 confirm the importance of acceptor pocket shape as defined by the size of amino acid side chains. All three UGTs possess the ability to glucosylate the flavonoid quercetin (Fig. 9). *Mt*UGT85H2 and *Vv*GT1 give rise to the formation of quercetin 3-*O*-glucoside. A second minor quercetin glucoside is also produced by *Mt*UGT85H2 (Li et al., 2007) while *Mt*UGT71G1 produces five different quercetin glucosides (Fig. 9) (He et al., 2006). Molecular docking studies of *Mt*UGT71G1 with quercetin show that quercetin can be positioned in different orientations in the active site and illustrate how each of the five hydroxy groups may be properly positioned for glucosylation. The 3OH and 3′OH groups (Fig. 9) give the best fitness scores. 3′OH glucosylation is favored by H-bond formation to S285 as well as to the sugar donor (Shao et al., 2005). It is therefore not surprising that the quercitin 3′-*O*-glucoside accounts for 70% of the total product formed. The crystal structure of *Vv*GT1 was solved with quercetin in the active site allowing for specific interactions to be studied. In *Vv*GT1, quercetin is positioned for 3OH glucosylation by H-bond formation to H150 similar interaction with kaempferol (Fig. 8E), Q84 and Q188 (Li et al., 2007; Offen et al., 2006). A comparison to *Mt*UGT85H2 shows that the Q84 residue occupies a slightly different position because it is situated in the differently positioned Nα3 (Fig. 8C) whereas the H150 and Q188 residues of *Vv*GT1 are replaced by two valine (V) residues in *Mt*UGT85H2. The valine residues do not offer the possibility of H-bond formation. This may explain why a differently glucosylated minor product is also formed (Li et al., 2007).

*Mt*UGT85H2 possesses a lower activity with quercetin compared to kaempferol as acceptor. In comparison to quercetin, kaempferol lacks the 3′OH group on the B-ring (Fig. 9). The activity differences may be explained by the presence of the large amino acid residue F202 pointing into the acceptor binding pocket and restricting space availability (Li et al., 2007). Both *Vv*GT1 and *Mt*UGT71G1 that show higher activity with quercetin compared to kaempferol have amino acid residues with small side chains at this position (Li et al., 2007). However, further analysis of the

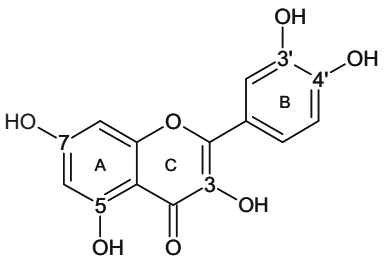


Fig. 9. Structure of Quercetin showing potential glycosylation sites. Glycosylation sites for the UGTs *Vv*GT1 (C-ring 3OH), *Mt*UGT85H2 (C-ring 3OH and unidentified site (Li et al., 2007)) and *Mt*UGT71G1 A-ring 5 and 7OH, C-ring 3OH and B-ring 3′ and 4′OH for *Mt*UGT71G1). In kaempferol the B-ring 3′OH is replaced by H.

residues lining the acceptor sites in the three crystal structures (loop N5a, Fig. 8F) show that the Y202 residue in *Mt*UGT71G1 may take up similar space in the 3D structure, even though in the primary sequence this residue is not positioned similarly to F202 in *Mt*UGT85H2. The importance of this residue for regioselectivity has actually been demonstrated by mutational studies in *Mt*UGT71G1. The point mutations Y202A and F148V, that both exchanges a large aromatic side chain with a smaller aliphatic side chain, switched the major glucosylation site from the 3′OH to the 3OH of quercetin (He et al., 2006) by allowing a different positioning of the acceptor. These results confirm the importance of the size and shape of the acceptor pocket for acceptor positioning and thus regiospecificity.

Successful changes in regioselectivity for the acceptor quercetin has also been demonstrated in mutational studies using *A. thaliana* UGT74F1 and UGT74F2 (Cartwright et al., 2008). One single amino acid substitution (N142Y) in *At*UGT74F1 changed the regiospecificity towards quercetin (Fig. 10). A secondary structure prediction of *At*UGT74F1 led us to include the sequence of this UGT in a structural alignment with the four plant UGTs with solved crystal structures. The residue N142 was predicted to be positioned within Nα5 that forms part of the acceptor pocket. The corresponding residues in *Vv*GT1, *Mt*UGT71G1, and *Mt*UGT85H2 depend on the structural alignment made, but were in the alignments we made here identified as histidine (H150), methionine (M152) and valine (V156), respectively (Fig. 10). The H150 from *Vv*GT1 is proposed to interact with the quercetin donor forming a stabilizing H-bond that keeps quercetin in place for 3OH glucosylation by *Vv*GT1 (Offen et al., 2006) (see above, and Fig. 8E). In the published structural alignment of the *At*UGT74F1 sequence with those of the four plant UGTs with solved crystal structures, the position of the N142 was displaced by two amino acids (Cartwright et al., 2008), compared to the alignment proposed here (Fig. 10). This would position N142 on the opposite side of Nα5, and prevent direct substrate interaction (Fig. 10). Clarification of this point must await a solved 3D structure for *At*UGT74F1.

In mutational studies of *Mt*UGT71G1 (He et al., 2006), independent substitutions of many other acceptor pocket residues into alanine residues resulted in decreased activity. The importance of space filling of a single amino acid residue for catalytic activity has also been reported for the *Catharanthus roseus* UGT2 (UGT73 family) (Masada et al., 2007). Based on these results, it can be concluded that the overall fit of the acceptor into the acceptor binding pocket as well as stabilization of the acceptor position by hydrophobic interactions are highly important for activity. This general conclusion is supported by the crystal structure of *At*UGT72B1

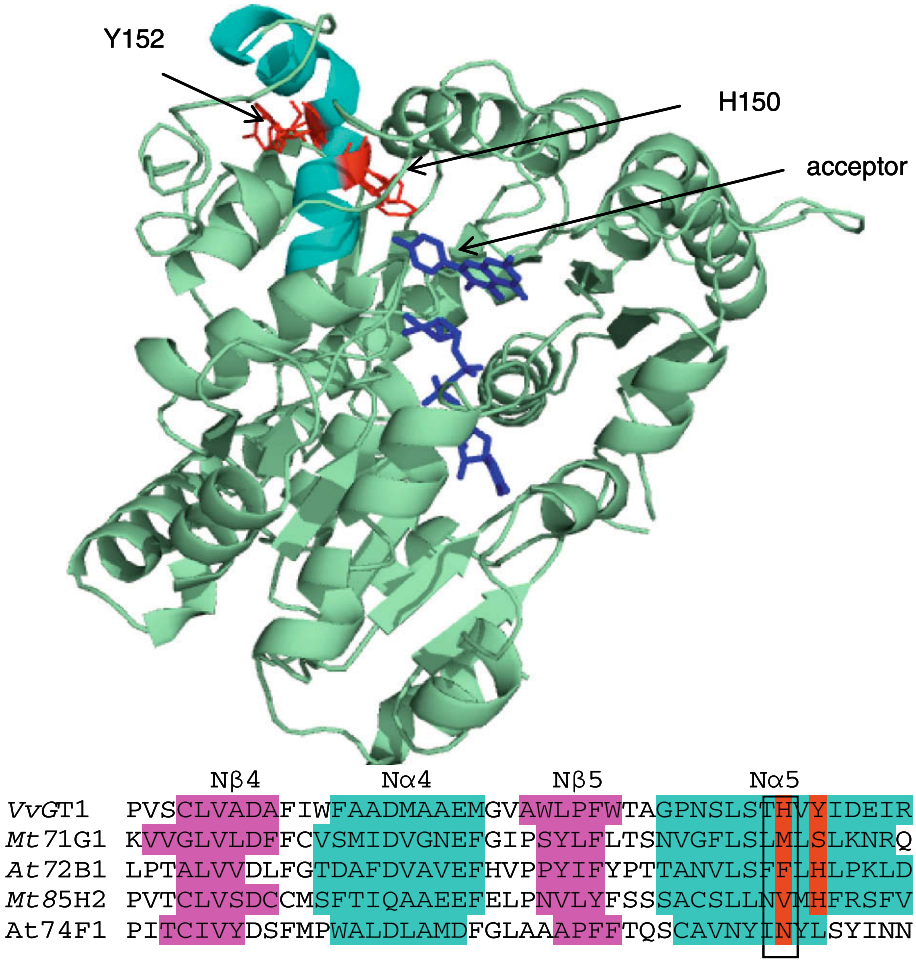


Fig. 10. Altered regioselectivity of *At*UGT74F1 towards quercetin glucosylation resulting from the point mutation N142Y (Cartwright et al., 2008). Panel A: The ribbon structure of the *Vv*GT1 crystal (Offen et al., 2006). The position of the *Vv*GT1 residues Y152 and H150 (shown in red) on either side of the N5α helix (cyan) is indicated. Substrates are shown as blue stick models. Panel B: Structural alignment of the region surrounding the *At*UGT74F1 amino acid residue N142 (boxed). The alignment includes sequences from *At*UGT74F1 and the four crystallized plant UGTs (Fig. 3A). Previous alignments (Cartwright et al., 2008) predicted residue Y152 from *Vv*GT1 to correspond to *At*UGT74F1 residue N141. The alignment performed here predicts the residue H150 from *Vv*GT1 to correspond to *At*UGT74F1 residue N141, i.e., a two amino acid shift as illustrated by the two columns of amino acids shown in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

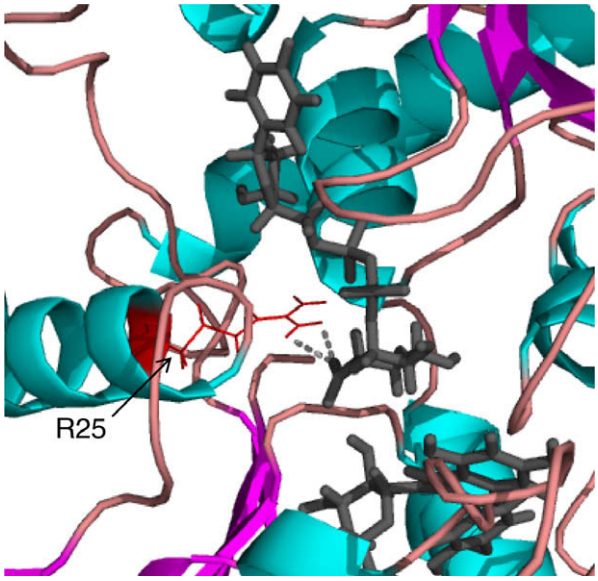


Fig. 11. Ribbon diagram of the homology model of the UDP-glucuronosyltransferase *Bp*UGT94B1. The side chain of R25 (shown as a red stick model) is predicted to be positioned favorably for interaction with the acidic group of the UDP-glucuronic acid sugar donor (indicated by dotted lines). α-Helices are shown in cyan, β-strands in pink and connecting loops in pale pink. The sugar donor and acceptor are shown as black stick models. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

which includes the acceptor TCP. The acceptor is positioned within the compact acceptor pocket by hydrophobic interactions alone as no hydrogen bonds can be formed between the acceptor and the amino acid residues lining the active site of the enzyme (Brazier-Hicks et al., 2007). Hydrophobic interactions are also observed to provide the majority of enzyme-acceptor interactions in other GT-B fold GTs (Bolam et al., 2007).

The region between Nβ5 and Nβ6 comprises the loops N5, N5a and N5b (Fig. 8E and F) and is the most diverging region not only among the plant UGTs but also when compared to and between other GT-B fold GTs. Available crystal structures of GT-B fold GTs suggest that differences in both the overall length and amino acid sequence of this region contribute to the large acceptor variety and specificity of these enzymes (Bolam et al., 2007; Ha et al., 2000; Mittler et al., 2007; Morera et al., 1999; Mulichak et al., 2001, 2003, 2004).

2.3.3. Inter- and intradomain interactions are important for activity and specificity

Stabilization of both secondary and tertiary structure can be conferred by intra- and interdomain interactions in the form of S–S bridges, salt bridges and H-bond formation. Interdomain interactions have been described in the *Mt*UGT85H2 crystal and include hydrogen bond formation between nine residue pairs each involving an N- and a C-terminal amino acid. This demonstrates potentially important interactions between the two domains despite the independent folding of these domains. Interestingly, of the nine interdomain interacting pairs described for *Mt*UGT85H2, seven include a residue within the PSPG motif. In a domain swap study involving interchange of the entire PSPG motif between *C. roseus* UGT2 (UGT73 family) and *Nicotiana tabacum* UGT71A7 (Masada et al., 2007), the chimeric enzymes were functionally inactive. This may reflect that unconserved residues within this otherwise highly conserved motif are engaged in intra- or interdomain interactions crucial for activity (Masada et al., 2007). The GT-B fold GtfA (Mulichak et al., 2003), GtfB (Mulichak et al., 2001) and GtfD (Mulichak et al., 2004) enzymes show marked differences in intra- and interdomain interactions. These interactions are thought to be important for the formation and stabilization of the acceptor binding site and to convey acceptor specificity (Mulichak et al., 2004).

Residues that are not predicted to be involved in any direct substrate interactions may influence acceptor specificity by their participation in intramolecular interactions as documented by analysis of the *At*UGT72B1 crystal structure. This UGT possesses dual *O*- and *N*-glucosylating activity. Mutation studies demonstrated that two residues, N312 and Y315, were determinants for *N*-glucosylation (Brazier-Hicks et al., 2007). These two residues are positioned in the C2 loop that is the most diverging of the C-terminal domain regions between UGTs with respect to overall length and amino acid sequence (Fig. 3A). The C2 loop is longer in *At*UGT72B1 compared to the other three crystal structures and is positioned closer to the acceptor. Nevertheless, it is too far for any direct interactions with the acceptor. The residue Y315 has been proposed to interact with P15 positioned within the N1 loop. This loop harbors the highly conserved and catalytically important H19 residue and the Y315–P15 interaction is proposed to position H19 in a position unfavorable for catalysis of *O*-glycosylation (Brazier-Hicks et al., 2007). In this manner the amino acid sequence of the C2 loop specifies *N*- versus *O*-glycosylation without any direct substrate interactions.

2.3.4. Domain swapping to obtain novel specificities of UGTs

The high structural conservation of UGTs combined with the observed variation in substrate specificity both for the sugar donor and acceptor implies that domain shuffling between UGTs possessing different substrate specificities could generate UGTs with novel acceptor and donor specificities. This principle may be taken a step further by exchanging the entire N- and C-terminal domains between UGTs. This intriguing possibility is based on the two domain structure of UGTs and assumes that donor and acceptor interactions across the domains in some cases may be neglected.

Several studies reporting domain swap between plant UGTs have been published (Cartwright et al., 2008; Kohara et al., 2007; Martin RC, 2000; Masada et al., 2007; Weis et al., 2008). Two of these studies were performed using *A. thaliana* UGTs and included domain swaps between UGTs of the same subfamily. Exchange of the entire N- and C-terminal domain between *At*UGT71C1 and *At*UGT71C3 sharing 81% similarity resulted in an active chimeric UGT. Km for the acceptor scopoletin was similar for both parental and chimeric proteins while the Km for the common sugar donor UDP-glucose matched that of the parental UGT providing the C-terminal domain. As for overall catalytic activity (k_{cat}) this was in between what was observed for the two parental enzymes (Weis et al., 2008).

Fusion of the N-terminal domain of *At*UGT74F2 and the C-terminal domain of *At*UGT74F1 have also provided an active chimeric protein (Cartwright et al., 2008). The parent enzymes share 90% sequence similarity and show similar Km values for the sugar donor UDP-glucose. In contrast, the two parent enzymes show differences in activity and regioselectivity with respect to the common sugar acceptor quercetin. The domain fusion protein showed a k_{cat} value and regiospecificity most like the UGT74F2 parental enzyme providing the N-terminal. Analyses of an additional six chimeric enzymes derived from the same parent enzymes by exchange of segments within domains also showed activity. These chimers allowed for identification of a single amino acid residue N142 as a determinant for regiospecificity, and enabled the construction of an enzyme with novel activity (Cartwright et al., 2008).

In a third set of domain swap experiments, domains and smaller segments were exchanged between UGT73L1 from *S. aculeatissimum* and *St*SGT (UGT73 family) from *S. tuberosum* sharing an amino acid sequence similarity of 75%. These two UGTs share the same specificity towards the sugar acceptor, but differ in their sugar

donor specificity. *St*SGT uses UDP-galactose and to a lesser degree UDP-glucose as sugar donor while *Sa*UGT73L1 is specific to UDP-glucose. Chimers constructed by swap of entire N- and C-terminal domains were inactive while five of seven segment swap chimers showed activity, but it was not possible to identify residues or regions that specified activity towards UDP-galactose (Kohara et al., 2007). This illustrates the complexity of defining the specificity of UGTs. In a domain swap experiment between two more distantly related UGTs, *C. roseus* UGT2 (UGT family 73) and *N. tabacum* UGT71A7 that share only 25% amino acid identity the chimeric enzymes obtained were also shown to be inactive (Masada et al., 2007). An assessment of the success rate in generation of active chimeric proteins with improved substrate specificities has recently been provided (Hansen et al., 2008). Domain swap studies of other family 1 UGTs from non-plant sources showed activity of chimers between N- and C-terminal domains of rat UDPGTr-3 and UDPGTr-4. The two parent enzymes share 85% amino acid sequence identity. In contrast, functional chimeric UGTs were not obtained by domain exchange between rat UGTs UDPGTr-2 and UDPGTr-4 that share only 65% sequence identity (Mackenzie, 1990). In a study of two *S. fradiae* family 1 GTs, UrdGT1a and UrdGT1c that shows 91% amino acid sequence identity, 9 of 10 constructed chimers were functionally active (Hoffmeister et al., 2001).

Domain swap studies thus show promise, as a way to identify single amino acids or peptide segments that define substrate specificity, and to gain an understanding of important intra- and interdomain interactions. In general, a successful domain swap appears to require a high amino acid sequence identity between parental UGTs, most likely to conserve essential intra- and interdomain interactions. This is in accordance with the described great complexity of UGT substrate recognition and activity, where several regions from both domains are involved. At the same time, the successful construction of active chimeric UGTs show, that as our understanding of the determinants of substrate specificity increases, engineering of functional chimers between more distant UGTs is likely to become achievable.

2.3.5. Random mutagenesis versus rational design for obtaining UGTs with novel substrate specificities

Directed evolution is an alternative approach to investigate which parts of an UGT sequence is involved in defining substrate specificity and to obtain UGTs with novel substrate specificities. This approach involves construction of UGT libraries from which mutated UGTs containing random combinations of single and multiple point mutations can be obtained. No such studies have been published with plant UGTs, but the approach has been used to create point mutants in the *S. antibioticus* UGT OleD. A screen of the mutant OleD library identified three OleD point mutations affording increased enzyme activity and less stringent substrate specificity (Williams et al., 2007). Combination of these three independent point mutations resulted in a triple mutant with synergistically increased glycosylation activity towards several acceptors with a range of different sugar donors not used by the wild type enzyme (Williams et al., 2007). Analysis of the crystal structure of OleD identified the three point mutated residues as potentially substrate interacting residues. However, our current knowledge is too sparse to deduce the substrate specificity properties of the mutants solely from analysis of the crystal structure. Random mutagenesis of UGTs would thus appear to constitute an alternative strategy to obtain UGTs with novel activities. Exploitation of the random mutagenesis approach is dependent on the availability of a robust high throughput screen that enables the formation of the glycoside of interest to be monitored. A unifying characteristic of a nucleotide sugar mediated glycosylation reaction is the accompanying release of a proton. Recently, a mutant library of the GT-A fold GT GTB from *H. Sapiens* was screened by colorimetric monitoring of the pH change in presence of the observed acceptor substrate or activated glycosidic donor (Persson and Palcic, 2008). This would appear to be an excellent general screen for GT activity.

2.4. Homology models of plant UGTs

The available crystal-based 3D structures of plant UGTs (Bolam et al., 2007; Li et al., 2007; Offen et al., 2006; Shao et al., 2005) have provided new insights with respect to inter- and intradomain interactions important for enzyme activity and specificity and have served to identify amino acid residues interacting with the sugar donor and acceptor. Unfortunately, only a limited number of UGT crystal structures have been solved why comparative modeling constitutes an attractive alternative for the study of 3D structures. Comparative modeling in the form of homology modeling relies on the conservation of secondary and tertiary structure between a query protein and a template protein for which the crystal-based 3D structure is available. Accordingly, the accuracy of homology modeling will steadily improve when more crystal templates become available (Chakravarty et al., 2005; Ginalski, 2006).

The two first homology-based model structures for plant UGTs were published before the crystal structure of any plant UGT had been solved. UGT73A5 from *D. bellidiformis* was modeled using coordinates from the crystal structure of the bacterial family 1 GT GtfB from *A. orientalis*. The amino acid sequence identity between the query protein and the template was as low as 14% (Hans et al., 2004). The homology model accurately predicted a number of catalytically important residues although it was subsequently realized that the sugar donor UDP-glucose was inserted in an inverted position. This resulted in wrong positioning of the acceptor. The incorrect position of substrates can be explained by the fact that the crystal structure of GtfB used as template neither contained bound sugar donor nor acceptor and thus did not provide specific information to guide their insertion. Accordingly, the sugar donor and acceptor substrates were inserted into to the model guided by the PSPG motif residues proposed to provide the sugar binding domain and by use of the docking software package Gold (Hans et al., 2004). The *Db*UGT73A5 model allowed the sugar donor pocket and many residues involved in UDP-sugar binding to be correctly predicted, while the exact interactions of these residues with the UDP-sugar donor were incorrect, in turn also leading to incorrect acceptor positioning. This shows the importance of having template structures co-crystallized with substrates or substrate analogs to accurately determine the amino acid residues that interact with the sugar donor and sugar acceptor. The study of Hans et al. also shows the limitations of automated docking. The available model enabled identification of substrate interactions involving the highly conserved key residues mainly situated within the PSPG motif, while the low sequence resemblance of *Db*UGT73A5 to the template made predictions of interacting residues positioned in loops and other less conserved regions difficult (Hans et al., 2004).

The second model of a plant UGT to be published was that of *Sorghum bicolor* UGT85B1 (Thorsoe et al., 2005) catalyzing the last step in synthesis of the cyanogenic glucoside dhurrin (Jones et al., 1999; Tattersall et al., 2001). At that time two additional crystal structures of family 1 GTs GtfA and GftD from *A. orientalis* had been solved. The model of UGT85B1 was built using GtfA and GtfB as templates. The amino acid sequence of the template proteins shared approximately 15% identity to that of *Sb*UGT85B1. An advantage of using the coordinates of the solved crystal structure of GtfA as template is that it also contained the sugar donor and acceptor. The known positions of the substrates within the crystal-based 3D structure of GtfA were used to guide substrate docking into the UGT85B1 model. The UGT85B1 model allowed for

accurate prediction of important residues interacting with the UDP-sugar donor. The predictive power of the model was verified by site directed mutagenesis of some of these residues (Thorsoe et al., 2005). As also experienced in the modeling of *Db*UGT73A5, difficulties were encountered in the modeling of loop regions and four loops, all positioned in the N-terminal domain, were not included in the model.

The models of *Db*UGT73A5 and *Sb*UGT85B1 verify the power of homology modeling for predicting important residues within structurally conserved regions of plant UGTs, even when identity in primary structure between the template and query sequence is as low as 15%. One of the major limitations of homology modeling is modeling of loop regions when using templates with low sequence conservation to the query protein (Chakravarty et al., 2005; Ginalski, 2006) and several loops were left out of the two models (Hans et al., 2004; Thorsoe et al., 2005). The GT crystals show that loop regions and the differences in these are important determinants of substrate specificity allowing for highly diverse acceptor molecules to be glycosylated. Availability of a larger number of crystal-based 3D structures of UGTs to improve modeling of loop regions is thus essential for the predictive power of homology modeling.

The first homology model of a plant UGT based on the coordinates of crystallized plant UGTs has recently been published. This study was made using the UDP-glucuronosyl transferase *Bp*UGT94B1 from *B. perennis* and crystal structure coordinates of *Mt*UGT71G1 and *Vv*GT1. The model included 416 of the 438 amino acids in *Bp*UGT94B1. The model enabled identification of residues involved in substrate specificity and activity of the enzyme. Mutational studies including residues predicted to be involved in substrate interactions verified the predictive power of the model. A single amino acid residue R25 was identified as the key determinant for specificity for the sugar donor UDP-glucuronic acid (Fig. 11) (Osmani et al., 2008).

2.5. The solved plant UGT crystals allow for improved homology modeling

In *A. thaliana*, phylogenetic studies have shown the PSPG containing plant UGTs to form a strong monophyletic clade (Paquette et al., 2003). Accordingly, in spite of the high structural conservation within GT-B fold GTs from one organism to the other, the amino acid sequence identity and structural homology is even higher within the PSPG containing plant UGTs. The availability of four solved plant UGT crystal structures thus offers better templates for homology modeling of plant UGTs compared to the crystal-based 3D structures of bacterial GTs formerly used. Future homology-based models of plant UGTs is therefore expected to be more accurate, especially with respect to the modeling of loop regions that show particularly low conservation to the bacterial GTs.

To assess the degree of homology model improvement resulting from the availability of crystal-based 3D structures of plant UGTs, we performed a re-modeling of the *S. bicolor* UGT85B1 using the coordinates of the two plant crystal structures of *Mt*UGT71C1 and *Vv*GT1 as templates. *Sb*UGT85B1 shares 26–30% amino acid identity with *Mt*UGT71C1 and *Vv*GT1. The model was built using the modeling software Sybyl from Tripos as also used for the former model (Thorsoe et al., 2005). A structural alignment of the new and old (Thorsoe et al., 2005) model is shown in Fig. 12. The alignment includes the sequence of GtfB that was one of the template structures used for making the old model. The new model based on the plant templates encompasses two of the four loops

Fig. 12. Structural alignment the old and new model of *Sb*UGT85B1 with the crystal structures of *Mt*UGT85H2 and GtfB. α-Helices are highlighted in cyan and β-strands in pink. Line 1: old *Sb*UGT85H1 model (Thorsoe et al., 2005); line 2: new *Sb*UGT85H1 model; line 3: crystal structure of *Mt*UGT85H2 sharing approximately 40% amino acid identity with *Sb*UGT85B1. NB: this structure was not used as template for the new model of *Sb*UGT85B1; line 4: crystal structure of the bacterial family 1 GT GtfB sharing approximately 15% amino acid identity with *Sb*UGT85B1. This structure was used as template for making the old *Sb*UGT85H1 model. Regions not clearly defined in the crystal structures or left out of the models are shown in bold face italics. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

that formerly could not be modeled and shortens the other two. In addition, a new loop is introduced at the interdomain linker, a new β-strand and three new α-helices are introduced while one α-helix is displaced and one shortened (Fig. 12). An overlay of the new and old UGT85B1 models (366 Cα's) afforded an RMSD value of 9.8 Å (Fig. 13, panel A). This signifies a quite large deviation between the two models. For comparison, the RMSD values of pairwise overlays between the four plant UGT crystals all ranged in the interval of 1.5–2.3 Å (368 Cα's).

The quality of the old and new model of *Sb*UGT85B1 (considering only Cα atoms) was assessed by calculation of their ProSA *Z*-scores (Wiederstein and Sippl, 2007). The *Z*-scores were −5.95 and −8.7, respectively, confirming improved quality of the new model. The *Z*-scores for experimentally determined X-ray structures of crystallized proteins of similar size (around 450 amino acids) lie in the range of −6 to −13 (Fig. 13, panel B). The *Z*-scores for the two template structures GtfA and GtfB used to build the old model were −11.95 and −9.85, respectively. *Z*-scores for the *Mt*UGT71G1 and *Vv*GT1 structures used as templates to build the new model were −12.75 and −10.39, respectively. Local model quality was calculated by the ProSA.residue score. The improved quality of the new model was apparent from the recording of fewer residues showing energy scores above 0, the threshold for indicating problematic regions (Fig. 1, panel C). Model quality was also determined by analysis using PROCHECK (Laskowski et al., 1992). This program has been designed to validate the stereochemical properties of all atoms in a protein structure by comparisons to well refined X-ray structures. The *G*-factors calculated by PROCHECK provide a measure of the probability of a given stereochemical property, a low *G*-factor indicates a low probability conformation. The overall *G*-factors were −0.93 and −0.36 for the old and new model, respectively. This indicates an improvement of the new model compared to the first one made, although both models ranged within the limits for good quality models (Bhattacharya et al., 2008). For comparison, the *G*-factors for the crystal-based 3D structures of *Vv*GT1 and *Mt*UGT71G1 used as templates were 0.13 and 0.24, respectively, and for the old model for GtfA (1PN3) and GtfB (1IIR) they were 0.29 and 0.43, respectively.

The quality parameters discussed above show that the new model of *Sb*UGT85B1 is likely to exhibit a higher resemblance with the crystal structure of *Sb*UGT85B1. The two models cannot be compared to the crystal-based 3D structure of *Sb*UGT85B1 as this has not yet been solved. After the new model of *Sb*UGT85B1 was completed, the crystal structure of a closely related UGT designated *Mt*UGT85H2 was published (Li et al., 2007). Because the two UGTs belong to the same family and share 40% amino acid sequence identity, this crystal structure offers an even better template for *Sb*UGT85B1 modeling. A comparison of the old and the new *Sb*UGT85B1 models to the *Mt*UGT85H2 crystal structure

A

RMSD (Å)	New model (466 Cα's)	New model N-terminal (203 Cα's)	New model C-terminal (230) Cα's)	New model PSPG motif (44 Cα's)
Old model	9.8 Å	4.8 Å	10.8 Å	2.0 Å

B

RMSD (Å)	*Mt*UGT85H2 (366 Cα's)	*Mt*UGT85H2 N-term (173 Cα's)	*Mt*UGT85H2 C-term (164 Cα's)	*Mt*UGT85H2 PSPG motif (44 Cα's)
New model *Sb*UGT85B1	3.4 Å	3.8 Å	1.4 Å	0.8 Å
Old model *Sb*UGT85B1	10.7 Å	12.6 Å	5.6 Å	1.6 Å

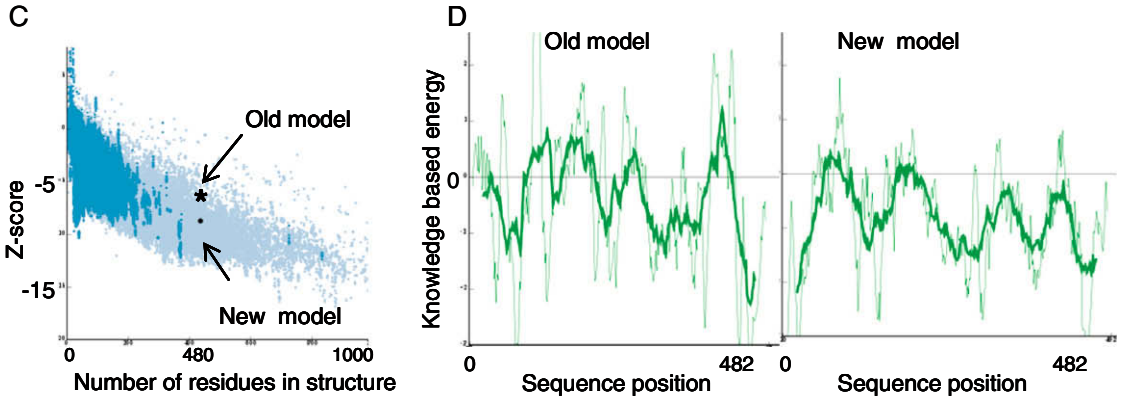


Fig. 13. A comparison of the quality of the new and old proposed structure of *Sb*UGT85B1 as obtained by homology modeling. Panel A: List of RMSD values calculated by overlay of the new and old models of *Sb*UGT85B1 and by overlay of the new and old models with the crystal structure coordinates of *Mt*UGT85H2. Panel B: The number of Cα's used in the different overlays are indicated. Panel C: ProSA evaluation of model quality showing the Z-score (*y*-axis) and the number of amino acids in the protein (*x*-axis). Dark blue zone shows *Z*-scores for experimentally determined NMR structures, Light blue zone shows *Z*-scores for experimentally determined crystal structures in the protein data bank. *Z*-scores for the new and old models are indicated. Panel D: ProSA residue score for the new and old model with amino acid number on the *x*-axis and energy score on the *y*-axis. Energy values are averaged over a window of 40 residues (dark green) and 10 residues (light green). Positive values indicate problematic regions. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

shows that the secondary structure of the new model offers the closest match (Fig. 12). Overlay of the crystal-based 3D structure of *Mt*UGT85H2 with the new and old model gives RMSD values (366 Cα's) of 3.4 Å and 10.7 Å, respectively. The RMSD values for the N-terminal domain (170 Cαs) are 3.8 Å and 12.6 Å, respectively, and those for the C-terminal domain (164 Cαs) are 1.4 Å and 5.7 Å, respectively (Fig. 14, panel A). These parameters show that the improvement of the new *Sb*UGT85B1 model is especially pronounced with respect to the more diverging N-terminal domain.

Overlays of the PSPG motif (44 Cα's) of *Mt*UGT85H2 with the new and old model of *Sb*UGT85B1 showed RMSD values of 1.6 Å and 2.0 Å, respectively. In accordance with these very good fits, visual inspection of the models confirmed the sugar donor interactions with PSPG motif residues predicted by the old *Sb*UGT85B1model, except for minor changes in the positions of some side chains. The old *Sb*UGT85B1 model predicted that the R201 residue positioned in the N5a loop interacted with the sugar donor (Thorsoe et al., 2005). This loop is differently positioned in the new model, and the previously proposed interaction was not validated. The new model is likely to be more accurate with respect to the R201 harboring region, as overlay of this region with the *Mt*UGT85H2 crystal structure showed a close fit. In the crystal structure of *Mt*UGT85H2, several residues in the R201 harboring region engage in interdomain interactions with PSPG motif residues (Li et al., 2007). These stabilizing interactions may explain why the *Sb*UGT85B1 R201A mutant show a 20-fold reduced activity (Thorsoe et al., 2005).

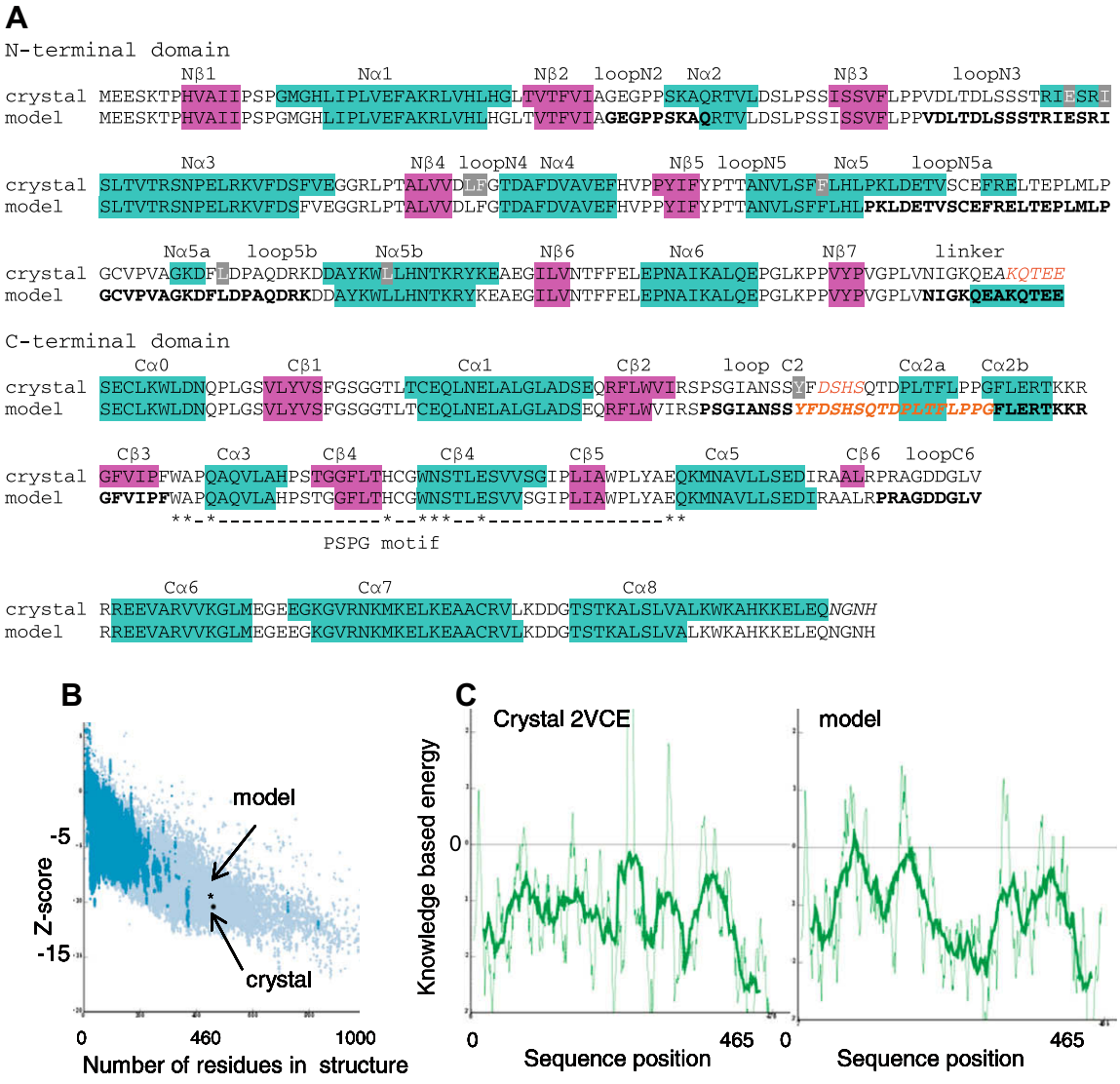


Fig. 14. Quality assessment of homology modeling based on a comparison of the crystal-based 3D structure of *At*UGT72B1 with a model of the same protein derived by homology modeling without use of the data for the crystal structure. Panel A: Alignment of the secondary structure of *At*UGT72B1 based on its crystal structure (RCSB protein data bank VCE) and derived from the structure obtained by homology modeling. The amino acid sequences shown in red are the regions missing from the crystal structure or homology model. Amino acid sequences shown in bold face mark the six most diverging regions between the crystal structure and the homology model (see text). The residue Y315 proposed to be involved in *N*-glycosylation by *At*UGT72B1 (Brazier-Hicks et al., 2007) is highlighted in red and the seven amino acid residues described to form the acceptor pocket of the crystal structure (Brazier-Hicks et al., 2007) are shown on a grey background (for their 3D position see Fig. 15B). Panel B: ProSA evaluation of the quality of the homology model showing the *Z*-score (*y*-axis) and the number of amino acids in the protein (*x*-axis), Dark blue zone shows *Z*-scores for experimentally determined NMR structures. Light blue zone shows *Z*-scores for experimentally determined crystal structures in the protein data bank. *Z*-scores for the homology model and crystal structures are indicated. Panel C: ProSA residue score for the crystal-based 3D structure and the structure derived by homology modeling with amino acid number on the *x*-axis and energy score on the *y*-axis. Energy values are averaged over a window of 40 residues (dark green) and 10 residues (light green). Positive values indicate problematic regions. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

In the present review, the discussion of the *Sb*UGT85B1 model has focused on identification of substrate interactions. However, *Sb*UGT85B1 is proposed to interact with the P450 proteins CYP79A1 and CYP71E1 (Bak et al., 1998; Sibbesen et al., 1995) to form a metabolon facilitating synthesis of the cyanogenic glucoside dhurrin (Jorgensen et al., 2005; Nielsen et al., 2008). Modeling of all three interacting partners with focus on surface exposed domains and loops with special functional characteristics such as hydrophobic or hydrophilic patches offers an approach to characterize the protein–protein interactions contributing to the formation of the dhurrin metabolon (DeLano, 2002).

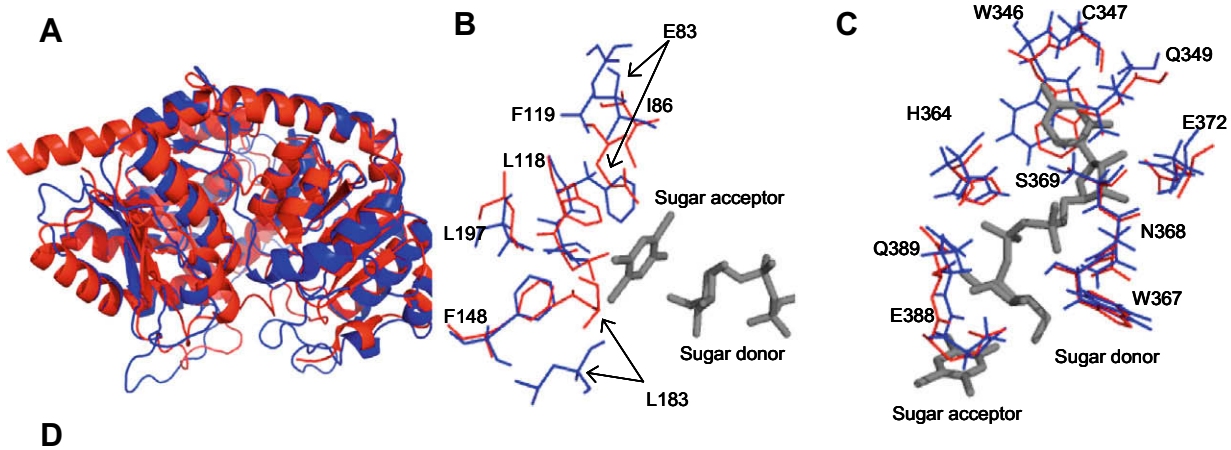
2.5.1. How good are UGT structures obtained by homology modeling?

It is apparent from the discussion above, that the short answer to this question is that this depends of the relatedness of the crystal structure template on which the UGT is to be modeled. The need to obtain 3D structures of many more UGTs to provide a good coverage of the many different families can hardly be overemphasized.

The most reliable test of how good homology modeling of UGTs performs compared to 3D structures obtained by crystallography would be to model a UGT for which the crystal structure is known without taking advantage of this knowledge. A subsequent direct comparison of the coordinates of the two structures is then possible. Modeling of *At*UGT72B1 was already in progress in our laboratory when the crystal structure of this UGT was published, providing an obvious possibility for such a study. The homology model was made using the coordinates of *Mt*UGT71G1 and *Vv*GT1 sharing 30–35% amino acid identity with *At*UGT72B1. The model comprises the amino acid residues 7–465 excluding the N-terminal (amino acids 1–7) and C-terminal (amino acids 465–480). Two loops could not be modeled from the templates. These were amino acid residues 243–257 corresponding to the interdomain linker region, and amino acid residues 304–325 corresponding to loop C2. The C2 loop is considerably longer in *At*UGT72B1 in comparison to the two template UGTs (Fig. 3A), which means that no template-based structure is proposed for this region. Instead, the two loop regions were modeled using other available protein crystal structures (excluding *At*UGT72B1) identified by a database search. The high uncertainty associated with modeling of these two regions and especially the 17 amino acid residue long C2 loop, was thus recognized from the beginning and should be kept in mind during assessment of the model (Fig. 14).

The quality of the structure derived from homology modeling was validated by calculating the ProSA *Z*-score (evaluating overall model quality by Cα positions) (Wiederstein and Sippl, 2007). The *Z*-score of the model was −10.39 while it was −11.43 for the crystal structure of *At*UGT72B1 (Fig. 14B). Local model quality calculated in ProSA also showed that the homology model had high quality with very few residues in problematic regions with energy scores above 0 (Fig. 14C). Model quality was also assessed by calculating the *G*-factors in PROCHECK (Laskowski et al., 1992). The overall *G*-factor evaluates all torsion angles and bond lengths of the model and was calculated to be −0.57, acceptable for a high quality model (Bhattacharya et al., 2008). For comparison, the crystal structure of *At*UGT72B1 (2VCE) have an overall *G*-factor of 0.07. PROCHECK analysis of the stereochemistry of individual residues can be used to identify problematic residues for further refinements of the model, but was not pursued here.

The crystal structure of *At*UGT72B1 does not include solved structures for the N- and C-terminal amino acid residues 1–6 and 471–480 as well as for the two regions containing amino acid residues 251–255 of the linker and residues 318–321 of the C2



D

RMSD (Å)	*At*UGT72B1 crystal 448 Cα's	*At*UGT72B1 crystal (428 Cα's) without C2 loop	*At*UGT72B1 crystal (317 Cα's) without 6 loops (see text)	*At*UGT72B1 crystal N-terminal 236Cα's	*At*UGT72B1 crystal C-terminal 189 Cα's (without C2 loop)	*At*UGT72B1 crystal PSPG motif (44Cα's)
*At*UGT72B1 model	3.6 Å	3.1 Å	1.8 Å	3..5 Å	1.9 Å	0.38 Å

Fig. 15. Comparing the model and crystal structure of *At*UGT72B1. Panel A: Ribbon diagram shoving an overlay of the crystal-based 3D structure *At*UGT72B1 (red) and the structure of the same protein obtained by homology modeling (blue). Panel B: Overlay of residues forming the acceptor pocket in the crystal structure (Brazier-Hicks et al., 2007) and homology model (red and blue stick models, respectively). The UDP-glucose and TCP are positioned based on the crystal structure and shown as black stick models. Panel C: Overlay of conserved PSPG residues interacting with the sugar donor in the crystal structure and homology model (red and blue stick models, respectively). Panel D: RMSD values of overlay of the *At*UGT72B1 crystal and homology model. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

loop. It can be noted that these two regions lie within the same two regions that could not be accurately modeled. A comparison of the secondary structure of the homology model and crystal shows high accordance between the two for regions containing elements of secondary structure (Fig. 14A). Of the 13 β-strands in the crystal, eleven are found in the model with conserved positions. One of the two missing β-strands is very short (2 amino acid residues) and the two residues are positioned similarly in the model and crystal. Of the 18 α-helices, all except N5α are present in the model, while differences in the length of Nα3 and Nα5 cause some differences between the crystal and model. Visual inspection of an overlay of the two structures identifies six regions that show small differences in position of the backbone between the model and the crystal. These are N2, N3, N5a–N5b, linker, C2–Cβ3 and C6. An overlay of the crystal and model (Fig. 15A) gives an RMSD of 3.6 Å for 448 Cα residues, and confirms the high agreement between the homology model and crystal. When the very tentative C2 loop is left out an overlay gives an RMSD of 3.1 Å (440 Cα residues and 3.9 Å for all atoms) while it is 3.5 Å and 1.9 Å for the N-terminal (Cα of amino acids 7–243) and C-terminal (Cα of amino acids 257–304 + 325–467) domains, respectively (Fig. 15D). Upon removal of the six loop regions that differs the most, an RMSD value of 1.8 Å for 317 Cα's is obtained. For homology modeled proteins sharing 30–50% amino acid sequence identity with their template, 80% of the Cα's in the model are expected to be within 3.8 Å of their correct position as determined by crystal structure (Kopp and Schwede, 2004). The homology model of *At*UGT72B1 was constructed from templates sharing 30–35% amino acid identity and 92% of the Cα's were within 3.1 Å of the positions in the crystal structure. This demonstrates a high accordance of the homology model with the determined crystal structure. RMSD values of 3.6 Å and 1.9 Å, obtained upon overlay of the N- and C-terminal domains of the homology model with the crystal-based 3D structure of *At*UGT72B1 reflect the expected higher conservation of the C-terminal domain as compared to the N-terminal domain between UGTs. RMSD values of 3.5 Å and 1.8 Å, respectively, were obtained upon comparison of an overlay of all Cα's and following deletion of the six most divergent loop regions (leaving 317 of the 465 amino acids in the model). This again illustrates the higher uncertainty associated with modeling of loop regions. In contrast, the high accuracy of the model for very conserved regions is apparent by overlay of the 44 amino acid PSPG motif of the homology model and crystal-based 3D structure which provided RMSD values as low as of 0.38 Å for all Cα's and 0.8 Å for all atoms (Fig. 15D).

Modeling of *At*UGT72B1 was completed by accommodation of the donor UDP-glucose and the acceptor TCP by homology docking (Imberty et al., 2006) guided by the substrate position in the crystal-based 3D structure of *Vv*GT1. The overlay shows that the sugar donor occupies the exact same position in the *At*UGT72B1 homology model and crystal structure. The acceptor is slightly displaced in the homology model. However, the acceptor O-atom is similarly positioned and the TCP ring is located in the same plane. A comparison of the conclusions made by analysis of the crystal structure with those obtained from the homology model illustrates the power of modeling as well as the shortcomings. All residues within the PSPG motif that are engaged in sugar acceptor interactions are positioned similarly in the model and crystal with respect to backbone as well as side chains (Fig. 15C). This is verified by the low RMSD value (Fig. 15D). In the crystal structure, the acceptor pocket is formed by the residues E83, I86, L118, F119, F148, L183 and L197 (Brazier-Hicks et al., 2007). A comparison of the position of these residues in an overlay of the crystal and the model shows that five of the seven residues are positioned nearly identically (RMSD values of 1.0 Å for Cα's and 2.0 Å for all atoms) (Fig. 15B). The two differently positioned residues are E83 and L183 that are found within the loops N3 (Nα3) and Nα5a, respectively. The loop N3 (Nα3) takes slightly different positions in the crystal structures of the plant UGTs (Fig. 8C). The *At*UGT72B1 crystal structure identifies the residues S16, H19 and D122 as directly involved in the catalytic mechanism (Brazier-Hicks et al., 2007). The side chain positions of these catalytically important residues shows slight differences between crystal and model (RMSD values of 0.6 Å for 3Cα's and 1.5 Å for all atoms of the three amino acid residues). The position of the three residues in the crystal in combination with residues in loop C2 offered an explanation for the ability of *At*UGT72B1 to carry out *O*- as well as *N*-glucosylations (Brazier-Hicks et al., 2007). The homology model does predict the involvement of the very same three amino acid residues in the catalytic reaction. However, the position of the side chain of H19 in the model cannot be used for a similar prediction. The positioning of the C2 loop region (residues 304–325) in the crystal brings the residue Y315 near to the acceptor site. A mutational study confirmed that this residue was important for catalytic activity (Brazier-Hicks et al., 2007). The C2 loop could not be accurately modeled, and the position of Y315 could therefore not be predicted. Part of this loop is also missing in the crystal structure and suggests that mobility of this loop is involved in its functionality.

2.6. Conclusions and perspectives: the determinants of plant UGT specificity

The mechanisms determining sugar donor and acceptor specificity and activity of UGTs are highly complex. The sugar acceptor and sugar donor substrates of plant UGTs are accommodated in the cleft formed between the N- and C-terminal domains. The amino acid sequences forming the sugar acceptor pocket are harbored within elements of conserved secondary structure and variable loop regions which may differ with respect to their amino acid sequence and sequence length. The crystal-based 3D structures of four plant UGTs have recently been published. They are all representatives of glucosyltransferases. To more precisely understand the determinants of UDP-sugar specificity, crystal structures of UGTs preferentially using other sugar donors than UDP-glucose would be of great interest. Likewise, availability of crystal structures representing plant UGTs with broad sugar acceptor specificity would also be of interest to give us a chance to understand how broad substrate specificity and high catalytic activity of such UGTs are accomplished. Availability of additional crystal-based 3D structures would greatly facilitate and improve structure prediction using homology modeling. In this review, the current predictive power of homology modeling was assessed to have a high degree of accuracy as demonstrated by comparison of a homology model of *At*UGT72B1 with its crystal-based 3D structure. Shortcomings in homology modeling are also apparent with modeling of loop regions remaining as a particularly difficult task. Conclusions based on structural studies should be supported by biochemical studies of the substrate specificity of wild type and mutated proteins. As the substrate specificity of more and more plant UGTs are being characterized, phylogenetic analysis may also gain importance in substrate prediction.

The number of UGT sequences is rapidly increasing, e.g., as a result of the EST and genome sequencing programs. Crystal-based 3D structure elucidation is a laborious process. NMR offers an attractive alternative to achieve structure elucidation of proteins in solution and to examine molecular dynamics. However, currently the masses of the UGTs are above the mass range amenable to NMR structural analyses. The C-terminal subunit of a family 1 GT Alg13 from *Saccharomyces cereviciae* was recently solved using NMR spectroscopy (Wang et al., 2008). The development of new techniques using gas-phase X-ray diffraction also offers new

avenues for structural studies of proteins and would allow studies using very small amounts of sample material (Hajdu, 2000). Advances within all these fields will contribute to our understanding of the relationship between amino acid sequence, 3D structure and substrate specificity of plant UGTs and facilitate the development of computer software to guide substrate specificity predictions.

References

Bak, S., Kahn, R.A., Nielsen, H.L., Moller, B.L., Halkier, B.A., 1998. Cloning of three A-type cytochromes P450, CYP71E1, CYP98, and CYP99 from *Sorghum bicolor* (L.) Moench by a PCR approach and identification by expression in *Escherichia coli* of CYP71E1 as a multifunctional cytochrome P450 in the biosynthesis of the cyanogenic glucoside dhurrin. Plant Mol. Biol. 36, 393–405.

Bhattacharya, A., Wunderlich, Z., Monleon, D., Tejero, R., Montelione, G.T., 2008. Assessing model accuracy using the homology modeling automatically software. Proteins 70, 105–118.

Bolam, D.N., Roberts, S., Proctor, M.R., Turkenburg, J.P., Dodson, E.J., Martinez-Fleites, C., Yang, M., Davis, B.G., Davies, G.J., Gilbert, H.J., 2007. The crystal structure of two macrolide glycosyltransferases provides a blueprint for host cell antibiotic immunity. Proc. Natl. Acad. Sci. USA 104, 5336–5341.

Bonina, F., Puglia, C., Rimoli, M.G., Melisi, D., Boatto, G., Nieddu, M., Calignano, A., La, R.G., De, C.P., 2003. Glycosyl derivatives of dopamine and L-dopa as anti-Parkinson prodrugs: synthesis, pharmacological activity and in vitro stability studies. J. Drug Target. 11, 25–36.

Bourne, Y., Henrissat, B., 2001. Glycoside hydrolases and glycosyltransferases: families and functional modules. Curr. Opin. Struct. Biol. 11, 593–600.

Bowles, D., Isayenkova, J., Lim, E.K., Poppenberger, B., 2005. Glycosyltransferases: managers of small molecules. Curr. Opin. Plant Biol. 8, 254–263.

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.

Breton, C., Snajdrova, L., Jeanneau, C., Koca, J., Imberty, A., 2006. Structures and mechanisms of glycosyltransferases. Glycobiology 16, 29R–37R.

Butelli, E., Titta, L., Giorgio, M., Mock, H.P., Matros, A., Peterek, S., Schijlen, E.G.W.M., Hall, R.D., Bovy, A.G., Luo, J., Martin, C., 2008. Enrichment of tomato fruit with health-promoting anthocyanins by expression of select transcription factors. Nat. Biotechnol. 26, 1301–1308.

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.

Caputi, L., Lim, E.K., Bowles, D.J., 2008. Discovery of new biocatalysts for the glycosylation of terpenoid scaffolds. Chemistry 14, 6656–6662.

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.

Chakravarty, S., Wang, L., Sanchez, R., 2005. Accuracy of structure-derived properties in simple comparative models of protein structures. Nucleic Acids Res. 33, 244–259.

Cheynier, V., 2005. Polyphenols in foods are more complex than often thought. Am. J. Clin. Nutr. 81, 223S–229S.

Cordel, G.A., 2002. Natural products in drug discovery. Phytochem. Rev. 1, 261–273.

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

DeLano, W.L., 2002. The PyMOL Molecular Graphics System. DeLano Scientific, Palo Alto, CA, USA.

Egleton, R.D., Mitchell, S.A., Huber, J.D., Janders, J., Stropova, D., Polt, R., Yamamura, H.I., Hruby, V.J., Davis, T.P., 2000. Improved bioavailability to the brain of glycosylated met-enkephalin analogs. Brain Res. 881, 37–46.

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.

Gachon, C.M., Langlois-Meurinne, M., Saindrenan, P., 2005. Plant secondary metabolism glycosyltransferases: the emerging functional analysis. Trends Plant Sci. 10, 542–549.

Geleijnse, J.M., Hollman, P.C., 2008. Flavonoids and cardiovascular health: which compounds, what mechanisms? Am. J. Clin. Nutr. 88, 12–13.

Gibson, R.P., Tarling, C.A., Roberts, S., Withers, S.G., Davies, G.J., 2004. The donor subsite of trehalose-6-phosphate synthase: binary complexes with UDP-glucose and UDP-2-deoxy-2-fluoro-glucose at 2 Å resolution. J. Biol. Chem. 279, 1950–1955.

Ginalski, K., 2006. Comparative modeling for protein structure prediction. Curr. Opin. Struct. Biol. 16, 172–177.

Ha, S., Walker, D., Shi, Y., Walker, S., 2000. The 1.9 A crystal structure of *Escherichia coli*, a membrane-associated glycosyltransferase involved in peptidoglycan biosynthesis. Protein Sci. 9, 1045–1052.

Hajdu, J., 2000. Single-molecule X-ray diffraction. Curr. Opin. Struct. Biol. 10, 569–573.

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, E.H., Osmani, S.A., Kristensen, C., Møller, B.L., Hansen, J., in press. Substrate specificities of family 1 UGTs gained by domain swapping. Phytochemistry, doi: 10.1016/j.phytochem.2009.01.013.

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

Harborne, J.B., Williams, C.A., 2000. Advances in flavonoid research since 1992. Phytochemistry 55, 481–504.

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

Hoffmeister, D., Ichinose, K., Bechthold, A., 2001. Two sequence elements of glycosyltransferases involved in urdamycin biosynthesis are responsible for substrate specificity and enzymatic activity. Chem. Biol. 8, 557–567.

Hoffmeister, D., Wilkinson, B., Foster, G., Sidebottom, P.J., Ichinose, K., Bechthold, A., 2002. Engineered urdamycin glycosyltransferases are broadened and altered in substrate specificity. Chem. Biol. 9, 287–295.

Hollman, P.C., Katan, M.B., 1997. Absorption, metabolism and health effects of dietary flavonoids in man. Biomed. Pharmacother. 51, 305–310.

Hou, B., Lim, E.K., Higgins, G.S., Bowles, D.J., 2004. *N*-glucosylation of cytokinins by glycosyltransferases of *Arabidopsis thaliana*. J. Biol. Chem. 279, 47822–47832.

Hu, Y., Chen, L., Ha, S., Gross, B., Falcone, B., Walker, D., Mokhtarzadeh, M., Walker, S., 2003. Crystal structure of the MurG: UDP-GlcNAc complex reveals common structural principles of a superfamily of glycosyltransferases. Proc. Natl. Acad. Sci. USA 100, 845–849.

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.

Imberty, A., Wimmerova, M., Koca, J., Breton, C., 2006. Molecular modeling of glycosyltransferases. Methods Mol. Biol. 347, 145–156. 145–156.

Jones, P., Messner, B., Nakajima, J., Schaffner, A.R., Saito, K., 2003. UGT73C6 and UGT78D1, glycosyltransferases involved in flavonol glycoside biosynthesis in *Arabidopsis thaliana*. J. Biol. Chem. 278, 43910–43918.

Jones, P.R., Moller, B.L., Hoj, 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.

Jones, P., Vogt, T., 2001. Glycosyltransferases in secondary plant metabolism: tranquilizers and stimulant controllers. Planta 213, 164–174.

Jorgensen, K., Rasmussen, A.V., Morant, M., Nielsen, A.H., Bjarnholt, N., Zagrobelny, M., Bak, S., Moller, B.L., 2005. Metabolon formation and metabolic channeling in the biosynthesis of plant natural products. Curr. Opin. Plant Biol. 8, 280–291.

Kamra, P., Gokhale, R.S., Mohanty, D., 2005. SEARCHGTr: a program for analysis of glycosyltransferases involved in glycosylation of secondary metabolites. Nucleic Acids Res. 33, W220–W225.

Kogawa, K., Kato, N., Kazuma, K., Noda, N., Suzuki, M., 2007. Purification and characterization of UDP-glucose: anthocyanin 3′,5′-*O*-glucosyltransferase from *Clitoria ternatea*. Planta 226, 1501–1509.

Kohara, A., Nakajima, C., Yoshida, S., Muranaka, T., 2007. Characterization and engineering of glycosyltransferases responsible for steroid saponin biosynthesis in Solanaceous plants. Phytochemistry 68, 478–486.

Kopp, J., Schwede, T., 2004. Automated protein structure homology modeling: a progress report. Pharmacogenomics 5, 405–416.

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

Kren, V., 2001. Glycosides in medicine: the role of glycosidic residue in biological activity. Curr. Med. Chem. 8, 1303–1328.

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.

Laskowski, R.A., MacArthur, M.W., Moss, D.S., Thornton, J.M., 1992. PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Cryst. 26, 283–291.

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.

Li, Y., Baldauf, S., Lim, E.K., Bowles, D.J., 2001. Phylogenetic analysis of the UDP-glycosyltransferase multigene family of *Arabidopsis thaliana*. J. Biol. Chem. 276, 4338–4343.

Lim, E.K., 2005. Plant glycosyltransferases: their potential as novel biocatalysts. Chemistry 11, 5486–5494.

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., 2003a. Evolution of substrate recognition across a multigene family of glycosyltransferases in *Arabidopsis*. Glycobiology 13, 139–145.

Lim, E.K., Bowles, D.J., 2004. A class of plant glycosyltransferases involved in cellular homeostasis. EMBO J. 23, 2915–2922.

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.

Lim, E.K., Higgins, G.S., Li, Y., Bowles, D.J., 2003b. Regioselectivity of glucosylation of caffeic acid by a UDP-glucose: glucosyltransferase is maintained *in planta*. Biochem. J. 373, 987–992.

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.

Mackenzie, P.I., Owens, I.S., Burchell, B., Bock, K.W., Bairoch, A., Belanger, A., Fournel-Gigleux, S., Green, M., Hum, D.W., Iyanagi, T., Lancet, D., Louisot, P., Magdalou, J., Chowdhury, J.R., Ritter, J.K., Schachter, H., Tephly, T.R., Tipton, K.F., Nebert, D.W., 1997. The UDP glycosyltransferase gene superfamily: recommended nomenclature update based on evolutionary divergence. Pharmacogenetics 7, 255–269.

Marcus, S.L., Polakowski, R., Seto, N.O., Leinala, E., Borisova, S., Blancher, A., Roubinet, F., Evans, S.V., Palcic, M.M., 2003. A single point mutation reverses the donor specificity of human blood group B-synthesizing galactosyltransferase. J. Biol. Chem. 278, 12403–12405.

Martin RC, C.K.M.M., 2000. Substrate specificity and domain analyses of zeatin O-glycosyltransferases. ISI Web of Science.

Masada, S., Terasaka, K., Mizukami, H., 2007. A single amino acid in the PSPG-box plays an important role in the catalytic function of CaUGT2 (Curcumin glucosyltransferase), a group D family 1 glucosyltransferase from *Catharanthus roseus*. FEBS Lett. 581, 2605–2610.

Mijatovic, T., Van, Q.E., Delest, B., Debeir, O., Darro, F., Kiss, R., 2007. Cardiotonic steroids on the road to anti-cancer therapy. Biochim. Biophys. Acta 1776, 32–57.

Miley, M.J., Zielinska, A.K., Keenan, J.E., Bratton, S.M., Radominska-Pandya, A., Redinbo, M.R., 2007. Crystal structure of the cofactor-binding domain of the human phase II drug-metabolism enzyme UDP-glucuronosyltransferase 2B7. J. Mol. Biol. 369, 498–511.

Miller, K.D., Guyon, V., Evans, J.N., Shuttleworth, W.A., Taylor, L.P., 1999. Purification, cloning, and heterologous expression of a catalytically efficient flavonol 3-*O*-galactosyltransferase expressed in the male gametophyte of *Petunia hybrida*. J. Biol. Chem. 274, 34011–34019.

Mittler, M., Bechthold, A., Schulz, G.E., 2007. Structure and action of the C–C bond-forming glycosyltransferase UrdGT2 involved in the biosynthesis of the antibiotic urdamycin. J. Mol. Biol. 372, 67–76.

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.

Morera, S., Imberty, A., Schke-Sonnenborn, U., Ruger, W., Freemont, P.S., 1999. T4 phage beta-glucosyltransferase: substrate binding and proposed catalytic mechanism. J. Mol. Biol. 292, 717–730.

Mulichak, A.M., Losey, H.C., Lu, W., Wawrzak, Z., Walsh, C.T., Garavito, R.M., 2003. Structure of the TDP-epi-vancosaminyltransferase GtfA from the chloroeremomycin biosynthetic pathway. Proc. Natl. Acad. Sci. USA 100, 9238–9243.

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.

Mulichak, A.M., Lu, W., Losey, H.C., Walsh, C.T., Garavito, R.M., 2004. Crystal structure of vancosaminyltransferase GtfD from the vancomycin biosynthetic pathway: interactions with acceptor and nucleotide ligands. Biochemistry 43, 5170–5180.

Nielsen, K.A., Tattersall, D.B., Jones, P.R., Moller, B.L., 2008. Metabolon formation in dhurrin biosynthesis. Phytochemistry 69, 88–98.

Offen, W., Martinez-Fleites, C., Yang, M., Kiat-Lim, E., 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, CA., Lindberg Møller, B., 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.

Ouzzine, M., Gulberti, S., Levoin, N., Netter, P., Magdalou, J., Fournel-Gigleux, S., 2002. The donor substrate specificity of the human beta 1,3-glucuronosyltransferase I toward UDP-glucuronic acid is determined by two crucial histidine and arginine residues. J. Biol. Chem. 277, 25439–25445.

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

Persson, M., Palcic, M.M., 2008. A high-throughput pH indicator assay for screening glycosyltransferase saturation mutagenesis libraries. Anal. Biochem. 378, 1–7.

Qasba, P.K., Ramakrishnan, B., Boeggeman, E., 2005. Substrate-induced conformational changes in glycosyltransferases. Trends Biochem. Sci. 30, 53–62.

Ramakrishnan, B., Boeggeman, E., Qasba, P.K., 2005. Mutation of arginine 228 to lysine enhances the glucosyltransferase activity of bovine beta-1,4-galactosyltransferase I. Biochemistry 44, 3202–3210.

Ramakrishnan, B., Qasba, P.K., 2002. Structure-based design of beta 1,4-galactosyltransferase I (beta 4Gal-T1) with equally efficient N-acetylgalactosaminyltransferase activity: point mutation broadens beta 4Gal-T1 donor specificity. J. Biol. Chem. 277, 20833–20839.

Rates, S.M., 2001. Plants as source of drugs. Toxicon 39, 603–613.

Ross, J., Li, Y., Lim, E.K., Bowles, D., 2001. Higher plant glycosyltransferases. Genome Biol. 2. reviews3004.1–reviews3004.6.

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 glucosyltransferase involved in flower pigment biosynthesis. J. Biol. Chem. 280, 899–906.

Sefton, M.A., Francis, I.L., Williams, P.J., 1994. Free and bound volatile secondary metabolites of *Vitis vinifera* grape cv. Sauvignon Blanc. J. Food Sci. 59, 142–147.

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.

Sibbesen, O., Koch, B., Halkier, B.A., Moller, B.L., 1995. Cytochrome P-450TYR is a multifunctional heme-thiolate enzyme catalyzing the conversion of L-tyrosine to p-hydroxyphenylacetaldehyde oxime in the biosynthesis of the cyanogenic glucoside dhurrin in *Sorghum bicolor* (L.) Moench. J. Biol. Chem. 270, 3506–3511.

Tanaka, Y., Yonekura, K., Fukuchi-Mizutani, M., Fukui, Y., Fujiwara, H., Ashikari, T., Kusumi, T., 1996. Molecular and biochemical characterization of three anthocyanin synthetic enzymes from *Gentiana triflora*. Plant Cell Physiol. 37, 711–716.

Tattersall, D.B., Bak, S., Jones, P.R., Olsen, C.E., Nielsen, J.K., Hansen, M.L., Hoj, P.B., Moller, B.L., 2001. Resistance to an herbivore through engineered cyanogenic glucoside synthesis. Science 293, 1826–1828.

Thorsoe, K.S., Bak, S., Olsen, C.E., Imberty, A., Breton, C., Moller, 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. Chembiochemistry 5, 16–25.

Thorson, J.S., Hosted, J.T., Jiang, J., 2001. Nature's carbohydrate chemists: the enzymatic glycosylation of bioactive bacterial metabolites. Curr. Org. Chem. 5, 139–167.

Tohge, T., Nishiyama, Y., Hirai, M.Y., Yano, M., Nakajima, J., Awazuhara, M., Inoue, E., Takahashi, H., Goodenowe, D.B., Kitayama, M., Noji, M., Yamazaki, M., Saito, K., 2005. Functional genomics by integrated analysis of metabolome and transcriptome of *Arabidopsis* plants over-expressing an MYB transcription factor. Plant J. 42, 218–235.

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

Vogt, T., 2002. Substrate specificity and sequence analysis define a polyphyletic origin of betanidin 5- and 6-O-glucosyltransferase from *Dorotheanthus bellidiformis*. Planta 214, 492–495.

Vogt, T., Grimm, R., Strack, D., 1999. Cloning and expression of a cDNA encoding betanidin 5-O-glucosyltransferase, a betanidin- and flavonoid-specific enzyme with high homology to inducible glucosyltransferases from the Solanaceae. Plant J. 19, 509–519.

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

Vogt, T., Zimmermann, E., Grimm, R., Meyer, M., Strack, D., 1997. Are the characteristics of betanidin glucosyltransferases from cell-suspension cultures of *Dorotheanthus bellidiformis* indicative of their phylogenetic relationship with flavonoid glucosyltransferases? Planta 203, 349–361.

Wang, X., Weldeghiorghis, T., Zhang, G., Imperiali, B., Prestegard, J.H., 2008. Solution structure of Alg13: the sugar donor subunit of a yeast N-acetylglucosamine transferase. Structure 16, 965–975.

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.

Weymouth-Wilson, A.C., 1997. The role of carbohydrates in biologically active natural products. Nat. Prod. Rep. 14, 99–110.

Wiederstein, M., Sippl, M.J., 2007. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res. 35, W407–W410.

Williams, G.J., Zhang, C., Thorson, J.S., 2007. Expanding the promiscuity of a natural-product glycosyltransferase by directed evolution. Nat. Chem. Biol. 3, 657–662.

Wriggers, W., Chakravarty, S., Jennings, P.A., 2005. Control of protein functional dynamics by peptide linkers. Biopolymers 80, 736–746.

Yonekura-Sakakibara, K., Tohge, T., Niida, R., Saito, K., 2007. Identification of a flavonol 7-*O*-rhamnosyltransferase gene determining flavonoid pattern in *Arabidopsis* by transcriptome coexpression analysis and reverse genetics. J. Biol. Chem. 282, 14932–14941.

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

Sarah A. Osmani obtained her MSc degree in 2003 from Department of Molecular Biology, University of Copenhagen where she studied stress induced protein expression in *Escherecia coli*. Her PhD studies have been carried out at The Plant Biochemistry Laboratory, Department of Plant Biology and Biotechnology at University of Copenhagen. During her PhD she isolated and characterized UDP-dependent glycosyl transferases (UGTs) from plants. The work involved molecular modeling of several plant UGTs as well as protein expression, site directed mutagenesis and biochemical studies. Studies on the color stability of anthocyanins has been an additional research topic. She is currently employed as a post-doctoral fellow in the Cyanogenic Glucosides and Molecular Evolution Research Group at The Plant Biochemistry Laboratory Department of Plant Biology and Biotechnology at University of Copenhagen working with modeling of metabolons incorporating cytochrome p450s and UGTs.

Søren Bak obtained his MSc in biochemistry at University of Copenhagen in 1993 and his PhD in Plant Molecular Biology at the Department of Plant Biology, Royal and Veterinary & Agricultural University (now faculty at the University of Copenhagen). The topic of the MSc thesis was within molecular plant virology. The topic of the PhD thesis was cytochromes P450 involved in primary and secondary metabolism in *Sorghum bicolor*. After his PhD thesis he initiated work on metabolic engineering of secondary metabolism in *Arabidopsis thaliana* and tobacco. During 1998–2000 he worked as a post-doctoral fellow at the Department of Plant Sciences, University of Arizona, US working on functional genomics of cytochromes P450 in *A. thaliana*. He is currently employed as Professor within systems biology and bioinformatics at the department of Plant Biology, University of Copenhagen where he is heading the Molecular Evolution Group. The long term goals of the research group are to elucidate the evolution of plant multigene families with focus on natural products and natural variation and their importance in plant–insect interactions. He is responsible for the Arabidopsis P450, Cytochromes b5, P450 Reductase and Glycosyltransferase site at http://www.p450.kvl.dk/.

Birger Lindberg Møller obtained his MSc, PhD and DSc from the University of Copenhagen in 1972, 1975 and 1984, respectively. His MSc Thesis work was focussed on identification of indole alkaloids in African medicinal plants. The topic of his PhD thesis was lysine metabolism in plants under supervision of Peder Olesen Larsen. He spent three years as a Fulbright Fellow at Eric Conńs laboratory, University of California, Davis, working with biosynthesis of cyanogenic glucosides. From 1977 to 1984 he was employed as Senior Scientist and Niels Bohr Fellow at the Department of Physiology, Carlsberg Laboratory with Diter von Wettstein and studying photosynthesis. This work formed the basis for his DSc thesis. In 1984 he was appointed Research Professor and in 1989 Professor in Plant Biochemistry at the Royal Veterinary and Agricultural University, now part of the University of Copenhagen. In 1998 he became the Head of Center for Molecular Plant Physiology (PlaCe) funded by two block grants from the Danish National Research Foundation for a 10-year period. In 2008 he has been appointed Head of the Villum Kann Rasmussen research centre "Pro-Active Plants". One of his main research interests is the synthesis, turn-over and storage of cyanogenic glucosides, the regulation of these processes and elucidation of the role of cyanogenic glucosides in plant insect and plant microbe interactions. The knowledge obtained is being used to ameliorate crop plants, fruit and forest trees like barley, sorghum, cassava, almonds, eucalyptus and aspen with respect to nutritive values and pest resistance. *Arabidopsis thaliana* and *Lotus japonicus* are used as model plants in these studies. Birger Lindberg Møller has been the main advisor of 33 PhD students working within these research areas. Birger Lindberg Møller is an elected member of the Royal Danish Academy of Science and Letters, of the Danish Academy of Technical Sciences, of the Danish Academy of Natural Sciences, of the Academic Council of the Faculty of Life Sciences, University of Copenhagen, Vice-Chairman of its Research and Innovation Board, member of the Board of Trustees of the International Institute of Tropical Agriculture (IITA) in Ibadan, Nigeria, Chairman of the Science Advisory Board at the Australian Research Council for appointment of Federation Fellows, member of the Advisory Board for the biotech company Aresa, and elected member of the International Human Rights Network of Academic and Scholarly Societies, Washington. In 2007, Birger Lindberg Møller was awarded the Villum Kann Rasmussen Research Prize, the largest Danish research award (350.000 Euro).