Functional characterization of a UDP-glucose:flavonoid 3-O-glucosyltransferase from the seed coat of black soybean (*Glycine max* (L.) Merr.)

Nik Kovinich [a,b,*], Ammar Saleem [c], John T. Arnason [c], Brian Miki [a]

[a] Bioproducts and Bioprocesses, Research Branch, Agriculture and Agri-Food Canada, Ottawa, ON, Canada K1A 0C6
[b] Ottawa–Carleton Institute of Biology, Department of Biology, Carleton University, Ottawa, ON, Canada K1S 5B6
[c] Department of Biology and Center for Research in Biopharmaceuticals and Biotechnology, University of Ottawa, Ottawa, ON, Canada K1N 6N5

ARTICLE INFO

ABSTRACT

The seed coats of black soybean (*Glycine max* (L.) Merr.) accumulate red (cyanidin-), blue (delphinidin-), purple (petunidin-), and orange (pelargonidin-based) anthocyanins almost exclusively as 3-O-glucosides; however, the responsible enzyme has not been identified. In this study, the full-length cDNA which encodes the enzyme that catalyzes the final step in anthocyanin biosynthesis, namely UDP-glucose:flavonoid 3-O-glucosyltransferase (*UGT78K1*), was isolated from the seed coat tissue of black soybean using rapid amplification of cDNA ends (RACE). Of the 28 flavonoid substrates tested, the purified recombinant protein glucosylated only anthocyanidins and flavonols, and demonstrated strict 3-OH regiospecificity. Galactose could also be transferred with relatively low activity to the 3-position of cyanidin or delphinidin *in vitro*. These findings are consistent with previous reports of mainly 3-O-glucosylated and minor amounts of 3-O-galactosylated anthocyanins in the seed coat of black soybean. The recombinant enzyme exhibited pronounced substrate inhibition by cyanidin at 100 μM acceptor concentration. Transfer of *UGT78K1* into the Arabidopsis T-DNA mutant (*ugt78d2*) deficient in anthocyanidin and flavonol 3-O-glucosyltransferase activity, restored the accumulation of anthocyanins and flavonols, suggesting the *in vivo* function of the enzyme as a flavonoid 3-O-glucosyltransferase. Genomic and phylogenetic analyses suggest the existence of three additional soybean sequences with high similarity to *UGT78K1*. RT-PCR confirmed the co-expression of one of these genes (Glyma08g07130) with *UGT78K1* in the seed coat of black soybean, suggesting possible functional redundancies in anthocyanin biosynthesis in this tissue.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

Black soybean (*Glycine max* (L.) Merr.) accumulates relatively high amounts of cyanidin-, delphinidin-, petunidin-, and pelargonidin-based anthocyanins exclusively as 3-O-glucosides in the seed coat (Choung et al., 2001; Lee et al., 2009). Minor amounts of 3-O-galactosides of cyanidin and delphinidin, and 3-O-glucosides of peonidin and a cyanidin–catechin complex have also been reported from this tissue (Lee et al., 2009). By contrast, anthocyanins from Arabidopsis contain glucosyl moieties linked to the 3- and 5-positions of the cyanidin backbone that are decorated with *p*-coumaroyl, malonyl, and sinapoyl groups (Tohge et al., 2005). Several genes involved in the early steps of anthocyanin and general flavonoid biosynthesis in the seed coat of black soybean have been identified, but the gene encoding the final step in anthocyanin biosynthesis, namely UDP-glucose:flavonoid 3-O-glucosyltransfer-

ase (UF3GT), has remained unreported (reviewed by Kovinich et al., 2011).

UF3GT catalyzes the transfer of glucose, from uridine diphosphate (UDP)-glucose, to the 3-position of anthocyanidins to form the corresponding anthocyanins (Forkmann and Heller, 1999; Heller and Forkmann, 1988, 1993). UF3GTs belong to a large multigene family (Family 1) of inverting glycosyltransferases (UGTs) (CAZy, http://www.cazy.org/fam/acc_GT.html) defined by the presence of a conserved carboxy-terminal consensus sequence, the plant secondary product glycosyltransferase signature sequence (PSPG box) (Hughes and Hughes, 1994), which is involved in binding the UDP moiety of the sugar nucleotide (Offen et al., 2006) to the enzyme. UGT Family 1 consists of over 100 members in Arabidopsis (Ross et al., 2001) and approximately 150 members in the legume *Medicago truncatula* (Modolo et al., 2007). Presently, neither UGT substrate specificity nor function can be predicted using amino acid sequence information alone (Modolo et al., 2007), and identity among UF3GTs has been shown to be as low as 25% (Sarkar et al., 2007). Some UGTs that catalyze the transfer of glucose to the 3-position of anthocyanidins can also glucosylate flavonol (Almeida et al., 2007; Ford et al., 1998; Ogata et al., 2004; Ralston et al., 1988; Tanaka et al., 1996), dihydroflavonol, flavone,

isoflavone (Owens and McIntosh, 2009), flavanone and coumestan substrates in vitro (Modolo et al., 2007). Identification and characterization of UGTs with high amino acid identities but different biochemical activities may provide insights into structure–function relationships and is the focus of ongoing studies (Osmani et al., 2009; Owens and McIntosh, 2009; Wang, 2009).

Anthocyanins are believed to be synthesized on the cytoplasmic face of the endoplasmic reticulum by metabolons (Saslowsky and Winkel-Shirley, 2001) and to be transported into the vacuole by systems that remain to be characterized (Grotewold and Davis, 2008). Glucosylation of the 3-position of anthocyanidins increases their chemical stability and glucosylation of secondary metabolites generally increases their solubility and enables access to active membrane transport systems that recognize glucosylated compounds but not their aglycones (Hostel, 1981). Glucosylation of the 3-position of anthocyanidins and the structurally similar flavonols may be required for their accumulation in some plant species, as a null mutation in the Arabidopsis UF3GT gene UGT78D2 (At5g17050) resulted in a drastic reduction of these compounds (79%) relative to the wild type (Tohge et al., 2005). The molecular basis of the black phenotype in soybean seed coats is not well understood, but the accumulation of high amounts of anthocyanins may be involved (Kovinich et al., 2011). Engineering reduced anthocyanin amounts in the seed coat of black soybean by suppression of the UF3GT gene could potentially be used to produce distinct seed colors to enable the visual identification and monitoring of transgenic grains (Kovinich et al., 2011). However, before engineering reduced anthocyanin amounts by suppression of the UF3GT gene can be attempted, the gene(s) encoding enzyme(s) catalyzing the final step in anthocyanin biosynthesis must be identified from the black soybean seed coat.

The purpose of this study was twofold: (1) to identify the gene that encodes the final step in anthocyanin biosynthesis from the seed coat of black soybean to provide a possibility of engineering reduced pigment in transgenic grains and (2) to characterize the catalytic properties of the recombinant enzyme to provide a basis for future glucosyltransferase structure–function analyses. We report herein the isolation of a glycosyltransferase cDNA (UGT78K1) from the seed coat of black soybean. To determine the function of the corresponding recombinant enzyme, 28 flavonoid substrates previously identified in soybean were tested as substrates. To determine the in vivo function, the cDNA was transferred into the Arabidopsis mutant ugt78d2 in an attempt to restore anthocyanin biosynthesis. To provide a basis for future structure–function analyses of UGT78K1, kinetics and specificity analyses were conducted with anthocyanidins and other flavonoid substrates.

2. Results

2.1. Cloning and analysis of a glycosyltransferase gene from the seed coat of black soybean

As the Glyma1 soybean genome sequence was not available at the time this study was initiated, G. max ESTs from the GenBank collection were searched with the tBLASTn and BLASTn algorithms using every published UF3GT sequence as a query (September, 2007), including the only published legume UF3GT (UGT78G1, from M. truncatula) (Modolo et al., 2007) in order to identify putative UF3GT candidates. However, these searches failed to identify any sequences with exceptionally high similarity. Highest similarity to GenBank G. max ESTs (up to 53%) was identified by tBLASTn using the PSPG box motifs of UGT78D2 from Arabidopsis (Tohge et al., 2005) and VvGT1 from Vitis vinifera (Ford et al., 1998). However, these searches yielded numerous unigenes with moderate similarity, and thus did not provide a sufficiently narrow pool of

candidates for functional analysis. The GenBank database was next searched using keywords such as anthocyanidin 3-O-glucosyltransferase which yielded the Ct3GT-A sequence (AB185904), representing a UF3GT from the legume Clitoria ternatea (Noda, personal communication). A megablast search using the Ct3GT-A mRNA as a query identified five ESTs with exceptionally high identity (68–81%). The ESTs were derived from various organs including seeds and flowers.

RNA was isolated from the seed coats of black soybean (G. max cv. Clark) and the three ESTs with highest identity to Ct3GT-A were extended by 5′- and 3′-RACE and sequenced. RACE fragments from the three ESTs were found to encode a single contiguous mRNA (1661 nts in length). The 5′- and 3′-untranslated regions were 86 bp and 234 bp, respectively. Conceptual translation of the mRNA gave a 447-amino acid protein, which was named UGT78K1. The presence of the plant secondary product glucosyltransferase (PSPG) consensus sequence in the C-terminal region of UGT78K1 identified it to be a type I glycosyltransferase (Vogt and Jones, 2000).

2.2. Genomic and phylogenetic analyses of soybean glycosyltransferases

To examine the potential for gene redundancy, a search for soybean genes with high identity to UGT78K1 (Glyma07g30180) was performed. A BLASTn search of the soybean (Glyma1) genome sequence (www.phytozome.net/soybean) afforded four sequences (Glyma07g30200, Glyma06g39350, Glyma07g30190, and Glyma08g07130) with high identity to UGT78K1 (83%, 85%, 86%, and 93% respectively). Glyma08g07130, Glyma07g30190, and Glyma07g30200 had high amino acid identity (93%, 79%, and 72% respectively) to UGT78K1; however, conceptual translation of Glyma06g39350 gave an highly truncated protein of 265 amino acids with two N-terminal deletions and the absence of the putative catalytic histidine residue equivalent to His20 of VvGT1 (Offen et al., 2006). This suggested that Glyma06g39350 may be non-functional, and thus it was excluded from further analysis.

Fig. 1 demonstrates the genomic organizations of UGT78K1 and the Arabidopsis gene UGT78D2 with their respective paralogs. Interestingly, the UGT78K1 genomic sequence is positioned in the same orientation and situated directly upstream from Glyma07g30190 and Glyma07g30200 (Fig. 1A), with a fourth highly similar sequence (Glyma08g07130) located on a separate
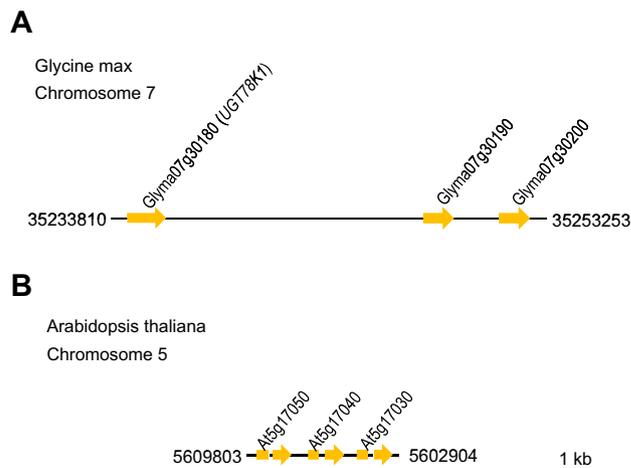

Fig. 1. Genomic organisations of UGT78K1 (Glyma07g30180), Glyma08g07130 and Glyma07g30190 from Glycine max (A) and Arabidopsis thaliana glycosyltransferases UGT78D2 (At5g17050), At5g17040, and At5g17030 (B).

chromosome (not shown). A similar genomic organization has been reported for *UGT78D2* (At5g17050, *ANL1*) and three paralogs (At5g17030, At5g17040, and At1g30530) in the Arabidopsis genome (Fig. 1B) (Kubo et al., 2007). However, unlike the Arabidopsis sequences, *UGT78K1* and its paralogous sequences lack the highly conserved glycosyltransferase (UGT) intron, equivalent to the Arabidopsis UGT intron 2 (Li et al., 2001). This data suggested that the gene triplication event that gave rise to *UGT78K1* and its paralogs on chromosome 7 (Glyma07g30190 and Glyma07g30200) most likely took place after divergence of *G. max* and Arabidopsis from a common ancestor, and thus the similar genomic organizations likely do not represent synteny. Furthermore, as Glyma08g07130 has greatest similarity to *UGT78K1*, this suggests that the gene duplication event that gave rise to this sequence occurred more recently than the triplication event that resulted in *UGT78K1*, Glyma07g30190, and Glyma07g30200 on chromosome 7.

Fig. 2 demonstrates a phylogenetic analysis of the amino acid sequences of biochemically characterized flavonoid UGTs, although the *in vitro* substrate specificities and *in vivo* functions of flavonoid UGTs cannot be accurately predicted using amino acid sequences alone. However, phylogenetic analysis resulted in five clusters that can be generally grouped based on common *in vitro* substrate specificities. Enzymes in Cluster V have anthocyanin 5-*O*- and/or flavonol 7-*O*-UGT activities, Cluster IV glycosylates flavonol and isoflavone substrates, and Cluster III includes isoflavone 7-*O*- and anthocyanidin 3,5-*O*-GT activities. Cluster II GTs have multiple substrate specificities generally for chalcones, flavones, and flavonols, but not anthocyanidins. Cluster I is exclusively UF3GTs. UGT78K1 groups with Cluster I, as does its paralogs (Glyma08g07130, Glyma07g30190, and Glyma07g30200). UGT78K1 groups more closely with its paralogs located on chromosome 7 (Fig. 1A) separately from UGT78D2 and its paralogs (Fig. 2). This further suggests that the gene triplication event that resulted in similar genomic organizations for *UGT78K1* and *UGT78D2*, and

their paralogs (Fig. 1), likely occurred after speciation and is not a result of synteny.

2.3. Expression analysis of black soybean glycosyltransferases

To determine whether the expression of *UGT78K1* coincided spatially with anthocyanin and/or flavonol biosynthesis in black soybean, RT-PCR was performed on RNA isolated from individual organs (Fig. 3). Anthocyanins accumulate to visible levels in the seed coat, mature stem, and purple flower petals of black soybean cv. Clark. By contrast, flavonol glycosides have been reported to accumulate in all aerial organs of soybean (Graham, 1991). RT-PCR identified *UGT78K1* transcripts in the seed coat, flower bud, stem, and all other organs examined (Fig. 3). These results suggested that *UGT78K1* is expressed not only at the sites of anthocyanin biosynthesis, but also at sites of flavonol glycoside biosynthesis. A similar expression pattern was observed for *UGT78D2* (*ANL1*) in Arabidopsis (Kubo et al., 2007). RT-PCR demonstrated overlapping expression profiles of Glyma07g30200, Glyma07g30190, and Glyma08g07130 with *UGT78K1* in some organs, but only Glyma08g07130 was co-expressed with *UF3GT* in the seed coat (Fig. 3).

2.4. UGT78K1 encodes a UF3GT with strict specificity for anthocyanidins and flavonols in vitro

The open reading frame of UGT78K1 was cloned into the pET-14b N-terminal hexahistidine fusion tag expression vector (Novagen), transformed into BL21(DE3) pLysS *Escherichia coli* cells, and expressed at 16 °C for 24 h. Protein was extracted and his-tagged UGT78K1 (rUGT78K1) was purified from the soluble fraction by ion-metal-affinity chromatography (IMAC) using the Talon system (Clontech, Palo Alto, CA). Purified fractions obtained from a 250 ml *E. coli* culture contained 1.6–2.4 mg of fusion protein. Separation of
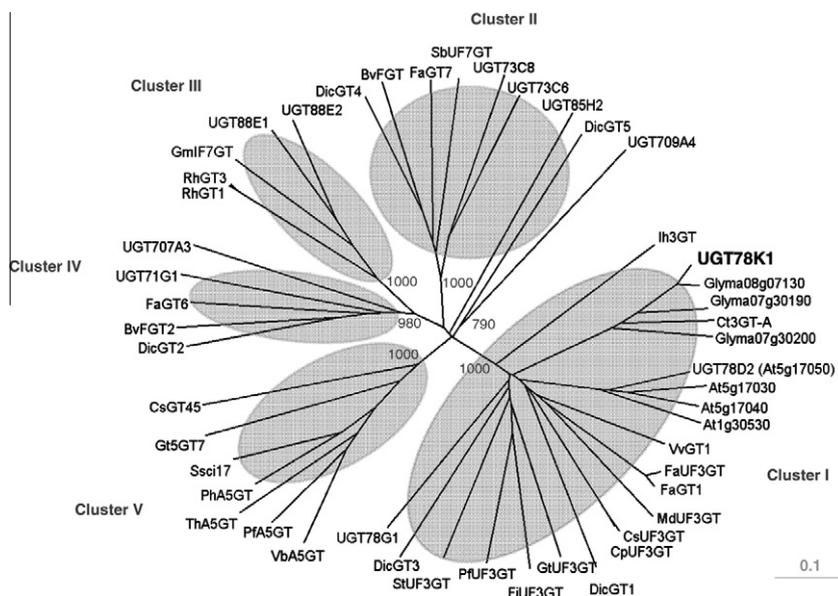

Fig. 2. Unrooted phylogenetic tree of UGT78K1 and flavonoid glycosyltransferase amino acid sequences. GenBank accessions: UGT78K1 (GU434274; Phytozome landmark: Glyma07g30180), Glyma07g30190 (Phytozome landmark: Glyma07g30190), Glyma07g30200 (Phytozome landmark: Glyma07g30200), Glyma08g07130 (Phytozome landmark: Glyma08g07130), Ct3GT-A (AB185904), UGT78D2 (NM_121711), At5g17030 (NC_003076), At5g17040 (NC_003076), At1g30530 (NM_102790), VhA5GT (AB013598), PfUA5GT (AB013596), GtUF3GT (D85186), VvGT1 (AF000372), Ih3GT (AB161175), DicGT1 (AB191245), DicGT2 (AB191246), DicGT3 (AB191247), DicGT4 (AB191248), DicGT5 (AB191249), RhGT1 (AB201048), RhGT2 (AB201049), RhGT3 (AB201050), GmIF7GT (AB292164), Ssci17 (AY033489), FaGT1 (AY663784), FaGT6 (DQ289587), FaGT7 (DQ289588), UGT73C8 (DQ875459), UGT88E1 (DQ875460), UGT88E2 (DQ875461), UGT71G1 (AAW56092), UGT85H2 (DQ875463), UGT78G1 (DQ875464), CsUF3GT (AY519364), FaUF3GT (AY695815), FiUF3GT (AF127218), MdUF3GT (DQ156906), PfUF3GT (AB002818), StUF3GT (AY954034), CpUF3GT (GQ141630), SbUF7GT (AB031274), BvFGT (AY526080), BvFGT2 (AY526081), UGT73C6 (NM_129234), UGT707A3 (BAC83989), UGT709A4 (BAC80066), Gt5GT7 (AB363839), CsGT45 (FJ194947). The number adjacent branches indicate maximum parsimony bootstrap values for the corresponding node. The scale bar indicates the number of differences per 100 residues derived from the ClustalW alignment. The phylogenetic tree was generated using TreeView software (Page 1996).
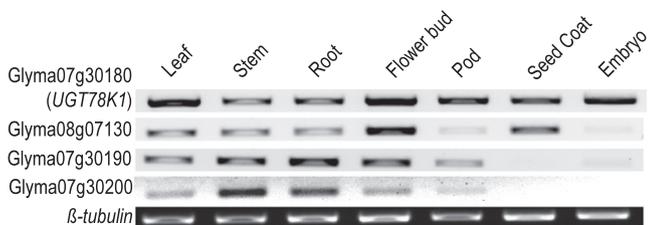
Fig. 3. Transcript detection of *UGT78K1* (Glyma07g30180), Glyma08g07130, Glyma07g30190, and Glyma07g30200 by RT-PCR with gene-specific primers.

the IMAC fractions by SDS–PAGE followed by staining with Coomassie Blue established an apparently pure recombinant enzyme with an estimated molecular mass of *ca.* 49.8 kDa (Fig. 4A). This value matched well with the calculated molecular mass of UGT78K1 (49.24 kDa) with the inclusion of its 2.18 kDa N-terminal tag. Preliminary assays using uridine diphosphate glucose (UDPG) as a sugar donor demonstrated that the purified rUGT78K1 could transfer glucose to the 3-position of the anthocyanidin cyanidin (**1**) (Fig. 4B and D) and the flavonol kaempferol (**3**) (Fig. 4C and E), whereas the boiled enzyme and bacteria expressing the corresponding empty vector could not catalyze these reactions (not shown). Glycosylated products (**2**) and (**4**) from the recombinant enzyme assays were identified and quantified by high performance liquid chroma-

tography-diode array detection (HPLC-DAD) by comparison to authentic standards. The presence of cyanidin-3-*O*-glucoside (C3G) (**2**) and kaempferol-3-*O*-glucoside (K3G) (**3**) reaction products were verified by high performance liquid chromatography electrospray ionization mass spectrometry (HPLC-ESI-MS) in the enhanced mass scan mode. The characteristic fragments detected in the C3G reaction mixture were 287.6 [cyanidin + H]$^{+1}$ and 450.1 [cyanidin-3-*O*-glucoside + H]$^{+1}$, and 287.1 [kaempferol + H]$^{+1}$ and 449.1 [kaempferol + H]$^{+1}$ for the K3G reaction mixture. Further purification of the enzyme by the removal of its hexahistidine tag did not alter enzyme activity, as determined by measuring activity before and after cleavage. For this reason, assays were routinely performed without removal of the tag. The enzyme exhibited no significant loss in activity towards kaempferol after storage at 4 °C or −20 °C in a 50% glycerol solution for 24 h; however, longer incubations resulted in a notable reduction in activity (not shown). To minimize the effect of storage on enzyme activity, assays were performed immediately following purification.

Optimal reaction conditions were determined using cyanidin (**1**) and kaempferol (**3**) as acceptor substrates with UDPG as the donor because cyanidin 3-*O*-glucoside (**2**) and kaempferol glycosides (e.g. (**4**)) are the most abundant anthocyanins and flavonols in the seed coats and purple petals of *G. max*, respectively (Choung et al., 2001; Iwashina et al., 2007; Lee et al., 2009). Investigation of a pH range from 5 to 10 and temperatures from 10 to 50 °C showed the
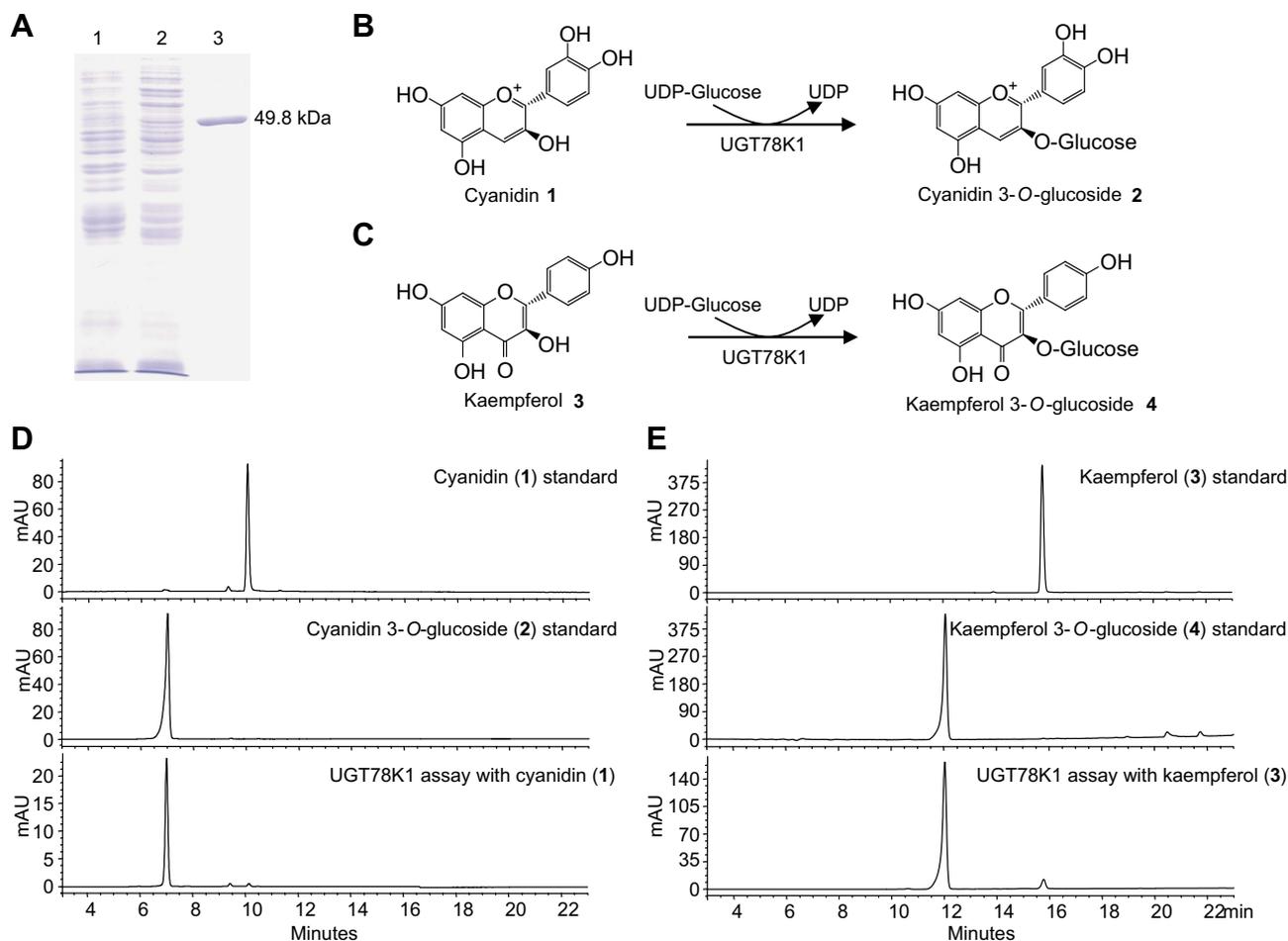

Fig. 4. Purification of his-tagged UGT78K1 and identification of reaction products. (A) SDS–PAGE analysis of UGT78K1: total soluble protein from *E. coli* expressing UGT78K1 prior to induction with 100 μM IPTG (lane 1), 24 h post-induction (lane 2); purified UGT78K1 (lane 3). (B, C) UGT78K1 enzyme reactions as revealed by HPLC chromatograms at 520 nm (D) and 350 nm (E), respectively. HPLC retention times: cyanidin (**1**) (*Rt*: 10.0 min); cyanidin 3-*O*-glucoside (**2**) (*Rt*: 7.0); kaempferol (**3**) (*Rt*: 15.8); kaempferol 3-*O*-glucoside (**4**) (*Rt*: 12.0).
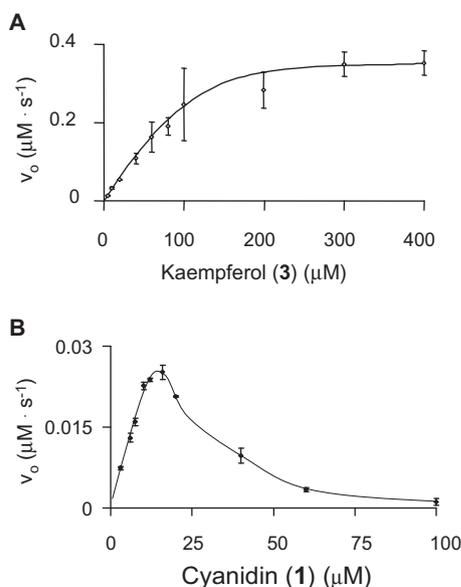
Fig. 5. UGT78K1 kinetics for cyanidin (**1**) and kaempferol (**3**) acceptor substrates. Reaction kinetics for acceptor substrates kaempferol (**3**) (A) and cyanidin (**1**) (B) using 5 mM UDPG as a sugar donor. Points represent the mean of three assays ± the standard deviation.

recombinant enzyme to be most active at pH 8 and 30 °C. rUGT78K1 exhibited a slight buffer preference for HEPES over Tris at pH 8 (not shown).

The kinetics of rUGT78K1 was determined for both cyanidin (**1**) and kaempferol (**3**) acceptor substrates (Fig. 5). By using a concentration of 5 mM UDPG and varying the concentration of kaempferol (**3**) between 5 and 400 μM, a hyperbolic saturation curve was obtained (Fig. 5A), from which Lineweaver–Burk transformation gave an apparent K_m for kaempferol (**3**) of 174 μM and V_{max} of 24.8 pKat μg^{-1}. From these a k_{cat} of 1.25 s^{-1} was calculated. By contrast, when the concentration of UDPG was maintained at 5 mM and the concentration of cyanidin (**1**) was varied from 5 to 100 μM, the reaction velocity increased with cyanidin (**1**) concentrations up to circa 16 μM and progressively decreased with increasing cyanidin concentrations (Fig. 5B), suggesting the enzyme was inhibited by this substrate.

The recombinant enzyme readily transferred glucose from UDPG to 3-position of all tested anthocyanidins and flavonol aglycones with variable specificities (Fig. 6). When acceptor substrates were supplied below the inhibitory concentration of cyanidin (**1**), the enzyme demonstrated highest activity towards the anthocyanidin delphinidin (Fig. 6A). By contrast, when acceptor substrates were supplied above the inhibitory concentration of cyanidin (**1**), UGT78K1 demonstrated highest activity for flavonols (quercetin, myricetin, and kaempferol) (**3**) (Fig. 6B, white bars). No activity was observed for other flavonoid classes previously identified in soybean, including dihydroflavonol (dihydroquercetin), isoflavone (genistin, genistein and daidzein), (epi)-flavan-3-ol (catechin, epicatechin), 3-deoxyanthocyanidin (apigeninidin), flavonol conjugate (flavonol 3-*O*-glucosides) and anthocyanins (anthocyanidin 3-*O*-glucosides and cyanidin 3-*O*-galactoside) (not shown). However, galactose could be transferred from UDP-galactose to the 3-position of cyanidin and delphinidin with low specificity relative to the analogous glucosyltransferase reactions (Fig. 6, grey bars), as determined by HPLC-DAD retention times and absorbance spectra by comparison to authentic standards. These results strongly suggest that rUGT78K1 acts exclusively on anthocyanidins and flavonols *in vitro* preferentially as a glucosyltransferase with strict regiospecificity for the 3-position.
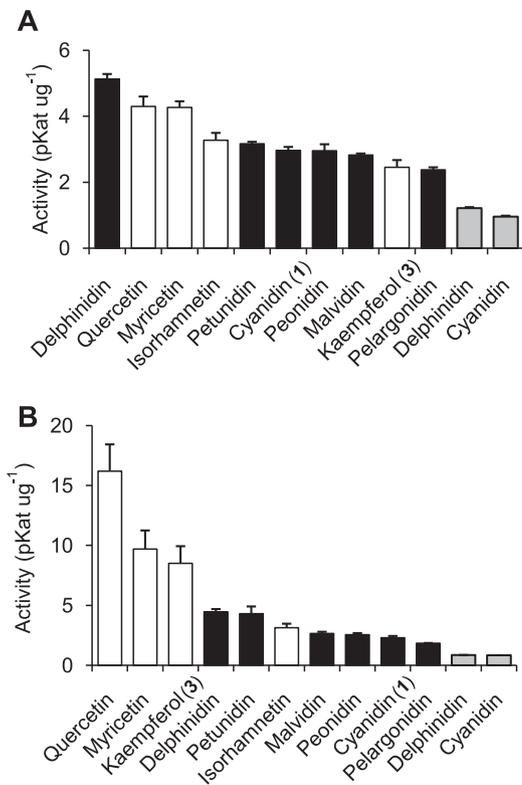

Fig. 6. Activities of UGT78K1 for anthocyanidins and flavonols. (A) Acceptor substrate concentration 10 μM. (B) Acceptor substrate concentration 100 μM. Anthocyanidins (black bars) and flavonols (white bars) assayed using 5 mM UDPG as a sugar donor. Cyanidin (**1**) and delphinidin assayed using 5 mM UDP-galactose as a donor (grey bars). Each bar represents the mean of three assays + the standard deviation.

2.5. UGT78K1 functions as a 3-O-glucosyltransferase in anthocyanin and flavonol biosynthesis in vivo

To confirm that *UGT78K1* has flavonoid 3-*O*-glucosyltransferase activity *in vivo*, the cDNA was transferred into the Arabidopsis mutant *ugt78d2* with highly reduced anthocyanidin and flavonol 3-*O*-glucosyltransferase activity (Tohge et al., 2005). T1 seedlings were selected on media containing kanamycin, transferred to soil and grown to maturity. Thirty percent of mature T1 plants (12 independent lines) selected on kanamycin media exhibited pigmentation at the rosette base similar to the wild type, whereas the *ugt78d2* mutant transformed with the empty vector lacked visible pigments (not shown). Three lines (U4-8, U4-25, and U4-28) were identified which carry a single copy of the *UGT78K1* cDNA by selection of T2 seedlings on kanamycin media ($\chi^2_{0.05,1}$ (3:1, viable:non-viable) = 1.7, 0.1, and 0.005, respectively) and were chosen for further analysis. Presence of the *UGT78K1* transgene and T-DNA insertions in the *UGT78D2* gene were verified by PCR using primers UHF/HUR and LB/URP, respectively, whereas the UGT78D2 was indentified to lack the T-DNA insertion using primers ULP/URP (Fig. 7A). RT-PCR was used to verify that each transgenic line expressed *UGT78K1* and lacked the expression of *UGT78D2* (Fig. 7B).

To examine the anthocyanin pigmentation phenotype more clearly, T2 seedlings cultured on anthocyanin induction media (see Section 5) and were visualized using light microscopy. The majority of T2 seedlings displayed highly pigmented cotyledons and hypocotyls similar to the wild type, whereas the *ugt78d2* mutant transformed with the empty vector had minimal amounts of visible pigments (Fig. 7C). To examine the amounts of total
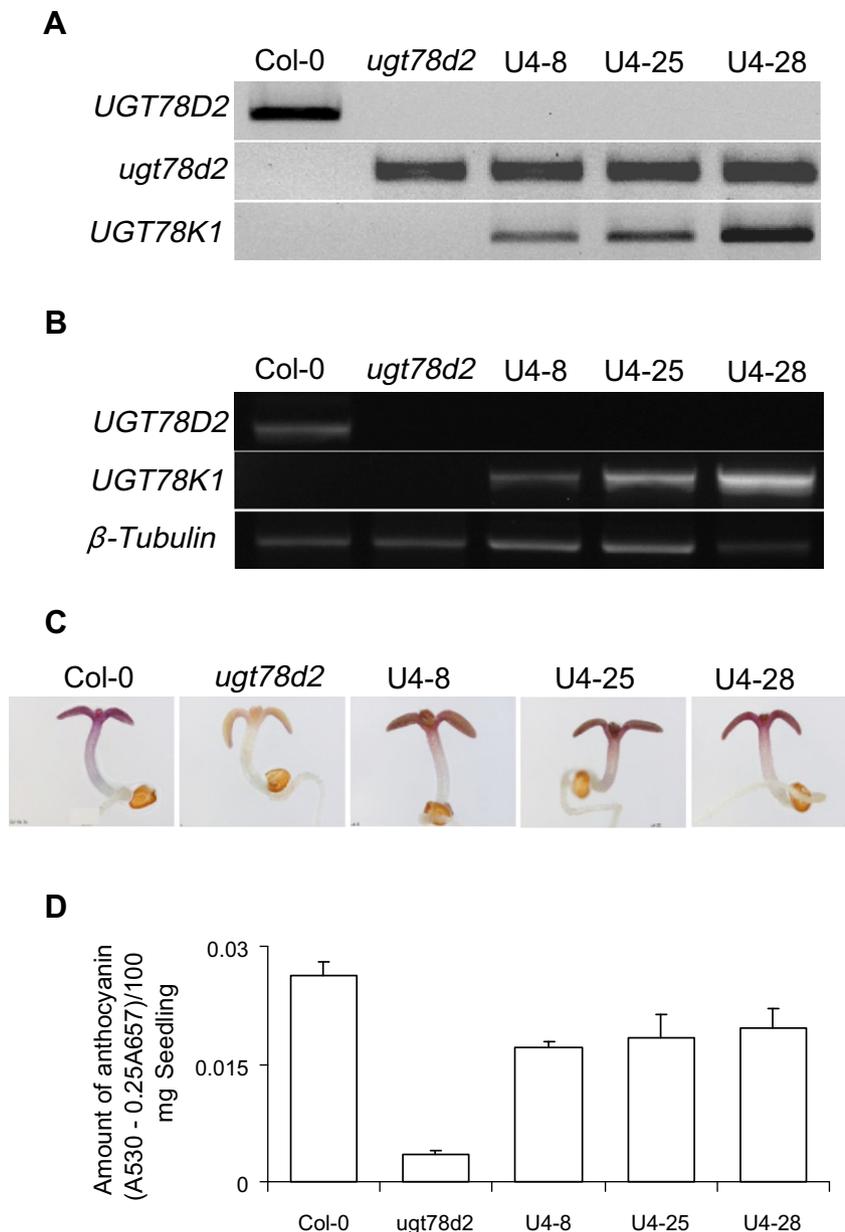
Fig. 7. The *UGT78K1* gene complements anthocyanin biosynthesis in the Arabidopsis *ugt78d2* mutant. (A) PCR analysis for the *UGT78K1* and *UGT78D2* genes and the T-DNA-inserted allele *ugt78d2* from genomic DNA of wild type (Columbia), mutant *ugt78d2* (transformed with empty pCAMBIA1300tCUP plasmid), and the *ugt78d2* mutant transformed with the *UGT78K1* gene (three independent T2 lines; U4-8, U4-25, U4-28). (B) RT-PCR analysis of total RNA from 100 mg of pooled seedlings from each line. (C) T2 lines cultured on anthocyanin induction media for 1 week (see Section 5). (D) Amount of anthocyanins in 100 mg of pooled T2 seedlings cultured on anthocyanin induction media.

anthocyanins in transgenic lines relative to wild type and *ugt78d2* controls, 100 mg of seedlings from each line were pooled, extracted with MeOH–H$_2$O (80:20), and total anthocyanins were quantified spectrophotometrically using a formula to compensate for chlorophyll absorption. On average, extracts from transgenic lines contained five (line U4-8) to six times (line U4-28) the anthocyanin absorbance relative to the *ugt78d2* mutant (Fig. 7D) and 65% (U4-8) to 75% (U4-28) absorbance relative to the wild type. This suggested restoration of anthocyanin biosynthesis in transformed seedlings, as 25% of the segregating T2s on non-selective anthocyanin induction media would not carry the *UGT78K1* gene.

To determine the relative amounts of individual anthocyanins and flavonols in transgenic lines relative to the wild type and

ugt78d2 controls, 100 mg of seedlings (T2s) from each line were pooled, extracted with MeOH–H$_2$O (80:20), and analyzed by HPLC-DAD. The *ugt78d2* mutant transformed with the empty vector had reduced peak area for all peaks measured at 520 nm (anthocyanins) and some peaks measured at 350 nm (flavonoids) relative to the wild type (Fig. 8). Highly similar profiles were reported previously for Col-0 and the *ugt78d2* mutant, and the peaks reduced at 520 nm and 350 nm in the *ugt78d2* mutant relative to the wild type were identified to be anthocyanins and flavonols, respectively, that contained a glucose at the 3-position (Tohge et al., 2005). The pooled T2 seedlings expressing the *UGT78K1* cDNA in the *ugt78d2* background had restored peaks, albeit to lower levels than the wild type. These results strongly suggest that
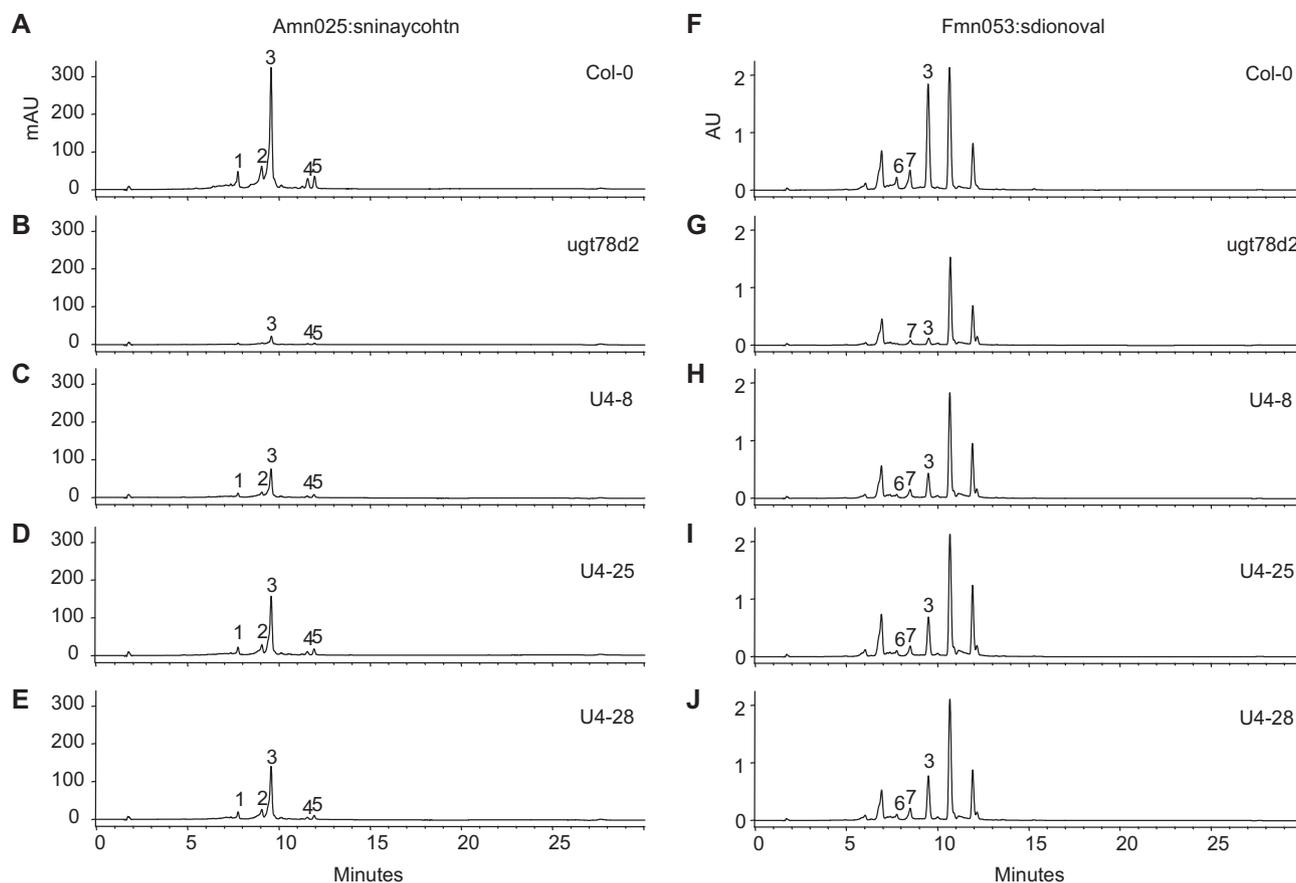
Fig. 8. Functional analysis of the *UGT78K1* gene by complementation of the Arabidopsis *ugt78d2* null mutation. (A–J) HPLC chromatograms of 80% aq. MeOH extracts of seedlings cultured on anthocyanin induction media for 1 week. (A, F) Col-0, (B, G) *ugt78d2*, (C, H) U4-8, (D, I) U4-25, (E, J) U4-28. (A–E) Absorbance at 520 nm for analysis of anthocyanins. (F–J) Absorbance at 350 nm for analysis of flavonoids. HPLC retention times: peak 1 (*R*t: 7.8); peak 2 (*R*t: 9.0); peak 3 (*R*t: 9.6); peak 4 (*R*t: 11.6); peak 5 (*R*t: 11.9); peak 6 (*R*t: 7.7); peak 7 (*R*t: 8.5).

UGT78K1 has similar activity to UGT78D2, a flavonoid 3-*O*-glucosyltransferase involved in anthocyanin and flavonol biosynthesis *in vivo*.

3. Discussion

Several genes involved in common flavonoid biosynthesis steps have been characterized from black soybean; however, the gene catalyzing the final step in anthocyanin biosynthesis, namely UDP-glucose:flavonoid 3-*O*-glucosyltransferase (*UF3GT*), remained unreported (reviewed by (Kovinich et al., 2011). The purpose of this study was twofold: (1) to identify the gene catalyzing the final step in anthocyanin biosynthesis from the seed coat of black soybean to provide a possibility of engineering reduced pigment in transgenic grains, and (2) to characterize the catalytic properties of the recombinant enzyme to provide a basis for future glucosyltransferase structure–function analyses.

The first UGT gene isolated from plants was the UDP-glucose:flavonol glucosyltransferase from *Zea mays*, encoded by the bronze locus (Fedoroff et al., 1984). Since then, UF3GT enzymes have been partially purified from *Brassica oleracea* (Saleh et al., 1976b), *Haplopappus gracilis* (Saleh et al., 1976a), *Senecio x hybridus* (Ogata et al., 1998), and cDNAs have been cloned into expression vectors for characterization of recombinant UF3GT activities from *Gentiana triflora* (Tanaka et al., 1996), *V. vinifera* (Ford et al., 1998), *Dianthus caryophyllus* (Ogata et al., 2004), *Iris hollandica* (Yoshihara et al., 2005), Arabidopsis (Tohge et al., 2005), *Fragaria*

x ananassa (Almeida et al., 2007), *M. truncatula* (Modolo et al., 2007; Peel et al., 2009) and *Citrus paradisi* (Owens and McIntosh, 2009). Homology-based searches of the GenBank database using each reported UF3GT as a query identified numerous soybean unigenes with equally low similarity. This suggested that soybean UF3GT(s) and other putative glycosyltransferases are highly divergent from published UF3GTs, including the only previously reported legume UF3GT UGT78G1 from *M. truncatula* (Peel et al., 2009).

All recombinant UF3GTs characterized to date have unique substrate specificities *in vitro*. Generally, some prefer anthocyanidins (Almeida et al., 2007; Ford et al., 1998; Ogata et al., 2004; Tanaka et al., 1996; Yoshihara et al., 2005), while others prefer flavonols and other (iso)flavonoids (Modolo et al., 2007; Owens and McIntosh, 2009). While UGT78K1 glucosylated only anthocyanidins and flavonols, its relative substrate preference was dependent on the concentration of acceptor substrate (Fig. 6). Kinetic analyses suggested the enzyme to be inhibited by cyanidin (**1**) at higher concentrations (above circa 16 μM) (Fig. 5B). The fact that high amounts of cyanidin (**1**) remained after enzyme reactions (not shown) supports the notion of substrate inhibition rather than its degradation as the cause of the observed kinetics. Future studies will include full kinetic analyses for all 10 acceptor substrates to determine whether they are similarly inhibiting UGT78K1 activity, and kinetic assays with an acceptor substrate (e.g. kaempferol (**3**)) and an inhibitor (cyanidin (**1**)) to gain more insight into the underlying mechanism of UGT78K1 inhibition.

UF3GTs from red cabbage (*B. oleracea*) and *H. gracilis* (Saleh et al., 1976a,b) have also been shown to be inhibited by cyanidin (**1**). However, it remains unclear whether inhibition of UF3GT activity by cyanidin could have a role in the regulation of flavonoid biosynthesis *in vivo*. The possibility exists that inhibition of UGT78K1 by cyanidin could prevent the cytotoxic effects of high cytosolic cyanidin 3-*O*-glucoside concentrations as observed in glutathione S-transferase mutants defective for vacuolar sequestration of anthocyanins (Alfenito et al., 1998; Marrs et al., 1995). Alternatively, inhibition of UGT78K1 by cyanidin 3-*O*-glucoside could function to make available more anthocyanidin substrates for proanthocyanidin biosynthesis, as anthocyanidins are precursors of both pathways (Lepiniec et al., 2006), and anthocyanins and PAs co-accumulate in palisade cells of black soybean during seed development (Kovinich, unpublished). However, another possibility is that localized cyanidin concentrations are not high enough in soybean to cause the substrate inhibition of UGT78K1 that is observed *in vitro*. Flavonoid enzymes are thought to physically associate in metabolons (Saslowsky and Winkel-Shirley, 2001; Winkel, 2009). Unfortunately, it is presently not possible to measure the localized concentrations of flavonoid intermediates within metabolons to determine whether these concentrations are comparable to those which are observed to cause inhibition *in vitro*.

Transfer of UGT78K1 into the Arabidopsis mutant *ugt78d2* restored both anthocyanin and flavonol glycoside biosyntheses (Fig. 8), confirming the *in vivo* function of UGT78K1 as a UF3GT. Anthocyanin levels were not restored to a lesser degree than flavonols in Arabidopsis despite the ability of cyanidin (**1**) to inhibit UGT78K1 *in vitro* (Fig. 5B). This suggested that substrate inhibition by cyanidin (**1**) was not operating *in vivo* and may have been a result of excessive amounts of the UGT78K1 enzyme relative to the amount of cyanidin (**1**) substrate. Similar to UGT78K1, the *M. truncatula* UF3GT UGT78G1 exhibited relatively low activity towards cyanidin (**1**) compared to other flavonoid substrates *in vitro* (Modolo et al., 2007), but was demonstrated to act in anthocyanin biosynthesis *in vivo* (Peel et al., 2009). These results raise the possibility that *in vitro* specificities may better reflect UF3GT capability rather than major *in vivo* function(s).

The V_{max} value of UGT78K1 for kaempferol (24.8 nKat μg^{-1}) was slightly higher than those reported previously for UF3GTs and flavonol substrates (11.6–20.5 nKat μg^{-1}) (Ford et al., 1998; Owens and McIntosh, 2009). The K_m values for flavonoid substrates from characterized UF3GTs vary greatly among species with values ranging from 1.5 to 400 μM (Ford et al., 1998; Modolo et al., 2007; Owens and McIntosh, 2009; Saleh et al., 1976a,b). The K_m of UGT78K1 for kaempferol (**3**) (174 μM) was within this range. Interestingly, UF3GTs that are inhibited by cyanidin (**1**) (174–400 μM) (Saleh et al., 1976a,b; the present study) have greater K_m values for flavonoid substrates than UF3GTs that do not exhibit substrate inhibition (12–90 μM) (Ford et al., 1998; Modolo et al., 2007; Owens and McIntosh, 2009). The identification of more UF3GTs and characterization of their kinetics for cyanidin (**1**) may help determine whether there is a relationship between K_m and inhibition.

Anthocyanins accumulate to visible levels in the seed coat, mature stem, and purple flower petals of black soybean cv. Clark. By contrast, flavonol glycosides have been reported to accumulate in all aerial organs of soybean (Graham, 1991) UGT78K1 was expressed transcriptionally in all organs, not just the sites of visible anthocyanin accumulations (Fig. 3) similar to UGT78D2 in Arabidopsis (Kubo et al., 2007). This expression pattern is similar to the UGT78D2 gene (At5g17050) in Arabidopsis (Kubo et al., 2007; Tohge et al., 2005) and may suggest that UGT78K1 also has a role in anthocyanin and flavonol glycoside biosyntheses *in vivo*.

UGT78K1 was demonstrated to catalyze the transfer of galactose to the 3-position of cyanidin (**1**) or delphinidin at low rates relative to the corresponding glucosyltransferase reactions (Fig. 6). These findings are consistent with previous reports of mainly 3-*O*-glucosylated and minor amounts of 3-*O*-galactosylated cyanidins and delphinidins in the seed coat of black soybean (Lee et al., 2009). However, knock-out or suppression of *UGT78K1* would be needed to determine whether this minor activity has a role the accumulation of anthocyanin galactosides in the seed coat of black soybean. We are currently attempting to obtain black soybean lines expressing an RNAi vector for silencing the UGT78K1 gene to test this hypothesis and to potentially change the color of the soybean seed coat.

4. Concluding remarks

In conclusion, we have isolated a cDNA from the seed coat of black soybean and have constructed a recombinant enzyme expression plasmid and transgenic Arabidopsis lines to determine the function of the encoded enzyme. Our results suggest that UGT78K1 is a UF3GT involved in anthocyanin and flavonol glycoside biosynthesis. Engineering reduced anthocyanin levels in the seed coat of black soybean by suppressing the expression of UGT78K1 could potentially produce distinct seed colors to enable the visual identification and monitoring of transgenic grains (Kovinich et al., in press). However, RT-PCR identified the co-expression of Glyma08g07130 (Fig. 3), which was demonstrated to be closely related to UGT78K1 by phylogenetic analyses (Fig. 2) and may suggest functional redundancy in this tissue. Future work will determine the acceptor substrate specificities of recombinant Glyma08g07130 for anthocyanidins to determine whether overlapping substrate specificities exist with UGT78K1. This will determine whether both genes must be silenced to reduce anthocyanin accumulations in the seed coat of black soybean.

5. Experimental

5.1. Chemicals

Cyanidin (**1**) was purchased from Indofine (Somerville, NJ, USA), delphinidin, myricetin, kaempferol, isorhamnetin, dihydroquercetin, apigeninidin, genistin, genistein and (−)-epicatechin were purchased from Sigma-Aldrich (Oakville, ON, CA), pelargonidin, malvidin, peonidin, kaempferol, quercetin, isorhamnetin and their 3-*O*-glucosides, petunidin, cyanidin 3-*O*-galactoside, (+)-catechin, apigenin from Extrasynthese (Lyon, FR) and 3-*O*-glucosides of delphinidin and petunidin from Polyphenols (Hanaveien, Norway).

5.2. HPLC-DAD and HPLC-ESI-MS analyses

The system (an Agilent 1100 series; Agilent Technologies Inc., Montreal, QC, Canada) was equipped with an autosampler with a 100 μL loop, a quaternary pump (maximum pressure, 400 bars), a column thermostat and a diode array detector (DAD). Separations were achieved at 45 °C on a Luna C18(2), 4.6 × 150 mm, particle size 5 μm fitted with a corresponding guard-column (Phenonenex Inc, Santa Ana, CA). The mobile phase system consisted of HPLC grade H_2O containing 5% HCO_2H (solvent A) and HPLC grade MeOH (solvent B). The optimized elution conditions were a linear gradient of 10–100% B in 25 min with a flow rate of 1 ml min^{-1}, the column was washed for 10 min at 100% B, brought back to starting mobile phase composition in 0.1 min and equilibrated for 5 min before next injection. The HPLC separations were monitored at 254, 280, 350, 476, and 520 nm.

HPLC-ESI-MS analysis used an 3200QTRAP (ABI Sciex, Toronto, Canada). The characteristic fragments of C3G and K3G were detected in enhanced mass scan (EMS) mode in positive ionization. The optimal settings of the mass spectrometer were: curtain gas (N_2) 10 L min^{-1}, ion spray voltage 4500 V, source gas (N_2) 10 L min^{-1}, declustering potential +50 V, entrance potential 10 V, ionization energy 1.0 eV and detector 2200.

5.3. Plant materials, growth conditions, and nucleic acid isolation

Black soybean (*G. max* (L.) Merr.) cv. Clark line (PI547438) was obtained from the U.S. Department of Agriculture Soybean Germplasm Collection (Agricultural Research Service, University of Illinois at Champaign–Urbana). Seeds were germinated in vermiculite and transferred to soil after one week. Plants were grown in a Conviron E15 cabinet under 16/8 h light/dark at 25 °C and 20 °C, respectively. Vegetative organs (obtained from 3 wko seedlings), seed coats, embryos and flower buds were dissected from plants, immediately frozen in liq. N_2, and stored at −80 °C. DNA was isolated from trifoliate either unexpanded soybean leaves or Arabidopsis seedlings using the DNeasy Plant Mini Kit (QIAGEN). RNA was isolated from all tissues using the RNeasy Plant Mini Kit (QIAGEN) or from black soybean seed coats as described previously (Wang and Vodkin, 1994).

Arabidopsis lines used in this study were of the Columbia ecotype. The T-DNA insertion line *ugt78d2* (CS25064) was obtained from the SALK collection (Alonso et al., 2003) at the ABRC (Columbus, OH, USA). Seeds were surface-sterilized in 2% triton/70% EtOH–H_2O (7:3, v/v) for 5 min on a mixer wheel, rinsed with EtOH (3×), dried, and plated on culture media (1X Gamborg's salts, pH 5.8, 1× B-5 vitamins, sucrose (2%, w/v), agar (0.8%, w/v) and 10 mg ml^{-1} benomyl). To release seed dormancy, plates were incubated for 3 d at 4 °C then transferred to light. Germination and seedling culture occurred in a Conviron I23L incubator (Conviron, Winnipeg, Canada) at 22 °C, under cool-white fluorescent lights (117 µE m^2 s) with a photoperiod of 16 h light to favor vegetative growth. Seedlings were harvested for nucleic acid isolation or transferred to soil after 2 weeks.

5.4. Cloning of the UF3GT cDNA from black soybean

By searching the NCBI *G. max* EST collection using the megablast algorithm and the *C. ternatea* Ct3GT-A mRNA (AB185904, Noda, Unpublished) as a query, forward and reverse gene-specific primers (B1F, B1R, B2F. B2R, E1F, E1R, M1F, M1R) (see Supplementary Table 1) were designed to amplify partial *UGT78K1* sequences from seed coat cDNA using 5′- and 3′-RACE (CLONTECH), respectively. The resulting partial cDNA fragments were cloned into the TOPO-TA vector (Invitrogen) and sequenced. The full-length coding region was then amplified from seed coat cDNA by end-to-end PCR using primers UHF/UHR (Supplementary Table 1) and the Pfx50 DNA polymerase (Invitrogen). The resulting 1517 bp amplicon was cloned into the *Nde*I and *Bam*HI sites of the pET-14b vector (Novagen) and sequenced to confirm its identity.

5.5. Cloning of the UF3GT gene from black soybean

The *UGT78K1* sequence was amplified from the soybean genome by PCR using primers UHF/UHR (Supplementary Table 1) and the Pfx50 DNA polymerase (Invitrogen), cloned into the pENTR/D-TOPO vector (Invitrogen), and sequenced. PCR of the wild type *UGT78D2* allele, the T-DNA insertion *ugt78d2*, and the *UGT78K1* gene was performed with primer pairs ULP/URP, LB/URP, and UNSF/UNSR, respectively (Supplementary Table 1).

5.6. Phylogenetic analyses

A multiple alignment of the deduced amino acid sequences of UGT78K1 and other flavonoid glycosyltransferases was constructed using ClustalW (http://www.align.genome.jp). Node support was estimated by parsimony bootstrap analysis (1000 bootstrap replicates, 10 random addition sequences per bootstrap replicate with tree bisection-reconnection branch swapping, limited to a maximum of 10,000 trees) using PAUP* 4.0b10 (Sinauer Associates).

5.7. RT-PCR

RNA samples (5 µg) were treated with DNase I (Amplification grade, Invitrogen) at 37 °C for 15 min prior to RT-PCR. First-strand cDNA synthesis was performed using SuperScript III Reverse Transcriptase (Invitrogen) according to the manufacturer's instructions. PCR of *UGT78D2*, *UGT78K1* (Glyma07g30190), Glyma08g07130, Glyma07g30190, Glyma07g30200, Arabidopsis β-tubulin, and *G. max* β-tubulin was performed with primers ULP/URP, UNSF/UNSR, 93F/93R, 86F/86R, 200F2/200R2, AtTubF/AtTubR, and GmTubF/GmTubR, respectively (Supplementary Table 1).

5.8. Recombinant protein expression in E. coli

The full-length *UGT78K1* coding region fused in-frame with the pET-14b vector (above) was transformed into the expression host *E. coli* BL21(DE3) pLysS (Novagen) and a single colony was selected for production of recombinant proteins. Soluble recombinant proteins were isolated from *E. coli* BL21(DE3) pLysS following growth and induction at 16 °C, and purified by ion metal-affinity chromatography as indicated (Ford et al., 1998). To confirm the purity of the recombinant enzyme the eluted fractions were visualized on 12.5% acrylamide gel stained with 0.25% Coomassie Blue. The amount of purified recombinant enzyme was determined using a protein assay kit (Bio-Rad Laboratories Inc.; http://www.bio-rad.com).

5.9. Recombinant enzyme assays

Initial enzyme assays (total volume 100 µl) consisted of recombinant enzyme (2 µg), acceptor substrate (100 µM) and 5 mM UDPG in assay solution (50 mM HEPES pH 8, 14 mM β-mercaptoEtOH) and were incubated at 30 °C for 30 min. To study the relationship between activity and pH, reactions were incubated in the following buffers, each at 50 mM: MES, pH 5–6; Tris, pH 7–9; HEPES, pH 7–8; CAPS, pH 10. To determine the effects of temperature on enzyme activity reactions were carried out in a thermocycler at 10–55 °C. To measure for linearity of product formation, reactions were stopped at 20 s intervals up to 3 min, using 5 mM UDP-glucose with 10 µM cyanidin (**1**) or 100 µM kaempferol (**3**).

The amount of each reaction product was determined by comparison of integrated peak areas of the glucoside (except for myricetin where the aglycone was used) to the corresponding standard curve analyzed on the same day. Standards were processed in the same manner as enzyme reactions with the exception that the recombinant enzyme was boiled prior to incubation. Standard curves were verified to be linear (COC > 98) over the range investigated using Chem32 software (Agilent Technologies Inc., Montreal, QC, Canada). All reaction solutions were brought to the specified assay temperature prior to the addition of the acceptor substrate and enzyme. Reactions containing either anthocyanins or 3-deoxyanthocyanidins were stopped by vortexing in HCO_2H–MeOH (500 µL, 15:85, v/v) for 20 s. Reactions were prepared for HPLC by centrifugation (21,000*g* for 4 min at 4 °C), followed by

removal of the supernatant and concentrated to 100 µL under a N$_2$ gas flow. All other reactions were stopped by vortexing in EtOAc (500 µl) for 20 s followed by removal of the organic phase, which was then evaporated under a stream of N$_2$. The residue was resuspended in HPLC grade MeOH (100 µl). All reactions were filtered through Teflon (0.2 µm; Chromatographic Specialties) and 20 µl aliquots were analyzed by HPLC-DAD.

5.10. Enzyme kinetics

To measure the reaction velocity of the recombinant enzyme for acceptor substrates, the concentrations of cyanidin (1) and kaempferol (3) were varied from 3 to 100 µM and from 5 to 500 µM, respectively, and 5 mM UDP-glucose was used as donor substrate. Enzyme assays were performed as described above but reactions were stopped at 3 min and 5 min ensure linearity of data points. Lineweaver–Burk graphs were plotted for the calculation of K_m and V_{max} values.

5.11. Enzyme specificity

The recombinant enzyme was tested for activity against anthocyaninidins (cyanidin (1), delphinidin, pelargonidin, petunidin, peonidin, malvidin), anthocyanins (3-O-glucosides of the anthocyanidins and cyanidin 3-O-galactoside), flavonols (kaempferol (3), quercetin, myricetin, isorhamnetin), flavonol-3-glucosides (3-O-glucosides of kaempferol, quercetin, and isorhamnetin), flavan-3-ols (+)-catechin and (−)-epicatechin, isoflavones (genistin, genistein and daidzein), a dihydroflavonol (dihydroquercetin), and a 3-deoxyanthocyanidin (apigeninidin), using 5 mM UDP-glucose as a donor substrate. UDP-galactose (5 mM) was also tested as a sugar donor using cyanidin and delphinidin acceptor substrates. Enzyme assays were performed as described above with either 10 or 100 µM of acceptor substrate and stopped at 3 min.

5.12. Transgenic lines

To obtain ugt78d2 lines transformed with UGT78K1, the full-length coding sequence of UGT78K1 was amplified from the pET-14b-UGT78K1 construct (described above) using primers BUF/BUR (Supplementary Table 1) and cloned into the BamHI and SacI sites of a modified pCAMBIA2300 vector (http://www.cambia.org) harboring the EntCUP4 constitutive promoter (Malik et al., 2002) and the NOS terminator between HindIII and EcoRI sites. Transformation of line ugt78d2 with the pCAMBIA2300-tCUP4::UGT78K1-tNOS construct was achieved using the Agrobacterium tumefaciens strain GV3101 as described (Clough and Bent, 1998). T1 seeds were selected on culture media and 50 µg ml^{-1} kanamycin then transferred to soil to set T2 seeds. Independent lines were identified to carry a single transgene copy if T2 seedlings exhibited a viable to non-viable ratio of 3:1 on selection media. Seedlings were visually screened for anthocyanin accumulation after 1 week of culture on anthocyanin gene induction media (pH 5.8) consisting of sucrose (3%, w/v), agar (0.8%, w/v) and 50 µg ml^{-1} kanamycin. Three independent lines which carried a single copy of the transgene, and which accumulated visible levels of anthocyanins, were selected for further analysis.

5.13. Metabolite analysis

One-week-old Arabidopsis seedlings (100 mg) cultured on anthocyanin gene induction media (above) were pulverized in MeOH–H$_2$O (200 µL, 80:20, v/v) in water using a FastPrep FP120 Homogenizer (Savant). The extracts were filtered through Teflon (0.2 µm; Chromatographic Specialties), with 20 µl aliquots analyzed by HPLC-DAD as described above, and 5 µl were measured by photospectroscopy using a NANODROP 2000 (Thermo Scientific) the formula A$_{530}$−0.25A$_{657}$ to compensate for chlorophyll absorption at 530 nm (Mancinelli, 1990).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.phytochem.2010.05.009.

References

Alfenito, M.R., Souer, E., Goodman, C.D., Buell, R., Mol, J., Koes, R., Walbot, V., 1998. Functional complementation of anthocyanin sequestration in the vacuole by widely divergent glutathione S-transferases. Plant Cell 10, 1135–1149.

Almeida, J.R., D'Amico, E., Preuss, A., Carbone, F., de Vos, C.H., Deiml, B., Mourgues, F., Perrotta, G., Fischer, T.C., Bovy, A.G., Martens, S., Rosati, C., 2007. Characterization of major enzymes and genes involved in flavonoid and proanthocyanidin biosynthesis during fruit development in strawberry (Fragaria xananassa). Arch. Biochem. Biophys. 465, 61–71.

Alonso, J.M., Stepanova, A.N., Leisse, T.J., Kim, C.J., Chen, H., Shinn, P., Stevenson, D.K., Zimmerman, J., Barajas, P., Cheuk, R., Gadrinab, C., Heller, C., Jeske, A., Koesema, E., Meyers, C.C., Parker, H., Prednis, L., Ansari, Y., Choy, N., Deen, H., Geralt, M., Hazari, N., Hom, E., Karnes, M., Mulholland, C., Ndubaku, R., Schmidt, I., Guzman, P., Aguilar-Henonin, L., Schmid, M., Weigel, D., Carter, D.E., Marchand, T., Risseeuw, E., Brogden, D., Zeko, A., Crosby, W.L., Berry, C.C., Ecker, J.R., 2003. Genome-wide insertional mutagenesis of Arabidopsis thaliana. Science 301, 653–657.

Choung, M., Baek, I.-Y., Kang, S.-T., Han, W.-Y., Shin, D.-C., Moon, H.-P., Kang, K.-H., 2001. Isolation and determination of anthocyanins in seed coats of black soybean (Glycine max (L.) Merr.). J. Agric. Food Chem. 49, 5848–5851.

Clough, S.J., Bent, A.F., 1998. Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. Plant J. 16, 735–743.

Fedoroff, N.V., Furtek, D.B., Nelson, O.E., 1984. Cloning of the bronze locus in maize by a simple and generalizable procedure using the transposable controlling element Activator (Ac). Proc. Natl. Acad. Sci. USA 81, 3825–3829.

Ford, C.M., Boss, P.K., Hoj, P.B., 1998. Cloning and characterization of Vitis vinifera UDP-glucose:flavonoid 3-O-glucosyltransferase, a homologue of the enzyme encoded by the maize Bronze-1 locus that may primarily serve to glycosylate anthocyanidins in vivo. J. Biol. Chem. 273, 9224–9233.

Forkmann, G., Heller, W., 1999. Biosynthesis of flavonoids. In: Sankawa, U. (Ed.), Comprehensive Natural Products Chemistry: Polypeptides and Other Secondary Metabolites Including Fatty Acids and Their Derivatives, vol. I. Elsevier Science Publishers, Amsterdam, pp. 713–748.

Graham, T.L., 1991. Flavonoid and isoflavonoid distribution in developing soybean seedling tissues and in seed and root exudates. Plant Physiol. 95, 594–603.

Grotewold, E., Davis, K., 2008. Trafficking and sequestration of anthocyanins. Nat. Prod. Commun. 3, 1251–1258.

Heller, W., Forkmann, G., 1988. Biosynthesis of flavonoids. In: Harborne, J.B. (Ed.), The Flavonoids: Advances in Research Since 1980. Chapman & Hall, London, pp. 399–425.

Heller, W., Forkmann, G., 1993. Biosynthesis of flavonoids. In: Harborne, J.B. (Ed.), The Flavonoids: Advances in Research Since 1986. Chapman & Hall, London, pp. 499–535.

Hostel, W., 1981. In: Stumpf, W., Conn, P.M. (Eds.), The Biochemistry of Plants, vol. 7. Academic Press, New York, pp. 725–753.

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.

Iwashina, T., Githiri, S.M., Benitez, E.R., Takemura, T., Kitajima, J., Takahashi, R., 2007. Analysis of flavonoids in flower petals of soybean near-isogenic lines for flower and pubescence color genes. J. Hered. 98, 250–257.

Kovinich, N., Arnason, J.T., De Luca, V., Miki, B., 2011. Coloring soybeans with anthocyanins? Recent Adv. Phytochem. in press, ISBN: 978-1-4419-6961-3.

Kubo, H., Nawa, N., Lupsea, S.A., 2007. Anthocyaninless1 gene of *Arabidopsis thaliana* encodes a UDP-glucose:flavonoid-3-O-glucosyltransferase. J. Plant Res. 120, 445–449.

Lee, J.H., Kang, N.S., Shin, S.-O., Shin, S.-H., Lim, S.-G., Suh, D.-Y., Baek, I.-Y., Park, K.-Y., Ha, T.J., 2009. Characterization of anthocyanins in the black soybean (*Glycine max* L.) by HPLC-DAD-ESI/MS analysis. Food Chem. 112, 226–231.

Lepiniec, L., Debeaujon, I., Routaboul, J.M., Baudry, A., Pourcel, L., Nesi, N., Caboche, M., 2006. Genetics and biochemistry of seed flavonoids. Annu. Rev. Plant Biol. 57, 405–430.

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.

Malik, K., Wu, K., Li, X.Q., Martin-Heller, T., Hu, M., Foster, E., Tian, L., Wang, C., Ward, K., Jordan, M., Brown, D., Gleddie, S., Simmonds, D., Zheng, S., Simmonds, J., Miki, B., 2002. A constitutive gene expression system derived from the tCUP cryptic promoter elements. Theor. Appl. Genet. 105, 505–514.

Mancinelli, A.L., 1990. Interaction between light quality and light quantity in the photoregulation of anthocyanin production. Plant Physiol. 92, 1191–1195.

Marrs, K.A., Alfenito, M.R., Lloyd, A.M., Walbot, V., 1995. A glutathione S-transferase involved in vacuolar transfer encoded by the maize gene Bronze-2. Nature 375, 397–400.

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.

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.

Ogata, J., Teramoto, S., Yoshitama, K., 1998. Isolation and characterization of UDP-glucose: cyanidin 3-O-glucosyltransferase from the flower buds of *Senecio x hybridus*. J. Plant Res. 111, 213–216.

Ogata, J., Itoh, Y., Ishida, M., Yoshida, H., Ozeki, Y., 2004. Cloning and heterologous expression of cDNAs encoding flavonoid glucosyltransferases from *Dianthus caryophyllus*. Plant Biotechnol. 21, 367–375.

Osmani, S.A., Bak, S., Moller, B.L., 2009. Substrate specificity of plant UDP-dependent glycosyltransferases predicted from crystal structures and homology modeling. Phytochemistry 70, 325–347.

Owens, D.K., McIntosh, C.A., 2009. Identification, recombinant expression, and biochemical characterization of a flavonol 3-O-glucosyltransferase clone from *Citrus paradisi*. Phytochemistry 70, 1382–1391.

Peel, G.J., Pang, Y., Modolo, L.V., Dixon, R.A., 2009. The LAP1 MYB transcription factor orchestrates anthocyanidin biosynthesis and glycosylation in Medicago. Plant J. 59, 136–149.

Ralston, E.J., English, J.J., Dooner, H.K., 1988. Sequence of three bronze alleles of maize and correlation with the genetic fine structure. Genetics 119, 185–197.

Ross, J., Li, Y., Lim, E., Bowles, D.J., 2001. Higher plant glycosyltransferases. Genom. Biol. 2, REVIEWS3004.

Saleh, N.A.M., Fritsch, H., Witkop, P., Grisebach, H., 1976a. UDP-glucose: cyanidin 3-O-glucosyltransferase from cell cultures of *Haplopappus gracilis*. Planta 133, 41–45.

Saleh, N.A.M., Poulton, J., Gris-Bach, H., 1976b. UDP-glucose: cyanidin 3-O-glucosyltransferase grom cabbage seedlings. Phytochemistry 15, 1865–1868.

Sarkar, T.R., Strong, C.L., Sibhatu, M.B., Pike, L.M., McIntosh, C.A., 2007. Cloning expression and characterization of a putative flavonoid glucosyltransferase from grapefruit (*Citrus paradisi*) leaves. In: Nikolau, B.J., Wurtele, E.S. (Eds.), Concepts in Plant Metabolomics. Springer, Dordrecht, pp. 247–257.

Saslowsky, D., Winkel-Shirley, B., 2001. Localization of flavonoid enzymes in Arabidopsis roots. Plant J 27, 37–48.

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.

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.

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

Wang, X., 2009. Structure, mechanism and engineering of plant natural product glycosyltransferases. FEBS Lett. 583, 3303–3309.

Wang, C.-S., Vodkin, L.O., 1994. Extraction of RNA from tissues containing high levels of procyanidins that bind RNA. Plant Molecul. Biol. Report. 12, 132–145.

Winkel, B.S.J., 2009. Metabolite channeling and multi-enzyme complexes. In: Osbourn, A.E., Lanzotti, V. (Eds.), Plant-derived natural products. Springer, New York, pp. 195–208.

Yoshihara, N., Imayama, T., Fukuchi-Mizutani, M., Okuhara, H., Tanaka, Y., Ino, I., Yabuya, T., 2005. CDNA cloning and characterization of UDP-glucose: anthocyanidin 3-O-glucosyltransferase in *Iris hollandica*. Plant Sci. 169, 496–501.