Phytochemistry 71 (2010) 1253–1263

Contents lists available at ScienceDirect

Phytochemistry

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

# 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

*Article history:*
Received 14 January 2010
Received in revised form 16 March 2010
Available online 4 June 2010

*Keywords:*
Black soybean
*Glycine max*
Leguminosae
*Arabidopsis thaliana*
Cruciferae
Glucosyltransferase
Galactosyltransferase
Anthocyanin
Flavonol
Seed coat
Pigment

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.



## 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*-glucosyltransferase (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,

* Corresponding author at: Bioproducts and Bioprocesses, Research Branch, Agriculture and Agri-Food Canada, Ottawa, ON, Canada K1A 0C6. Tel.: +1 613 715 5289; fax: +1 613 759 1701.
*E-mail address:* kovinichn@agr.gc.ca (N. Kovinich).

0031-9422/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.phytochem.2010.05.009

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

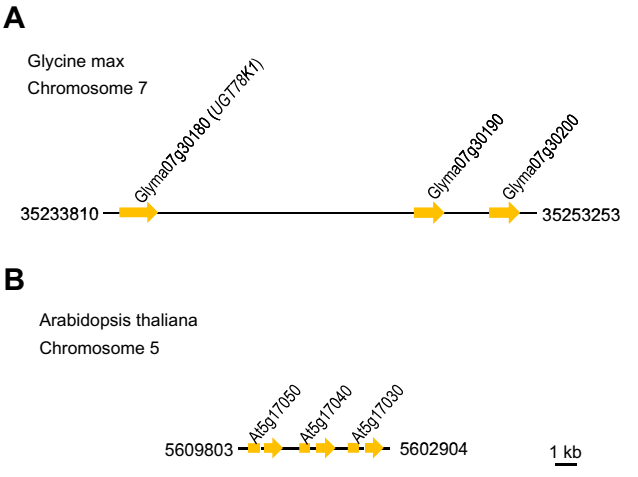


**Fig. 1.** Genomic organisations of *UGT78K1* (Glyma07g30180), Glyma08g07130 and Glyma07g30190 from *Glycine max* (A) and *Arabidopsis thaliana* glycosyltransferases UGT78D2 (At5g17050), At5g17040, and At5g17030 (B).