Phytochemistry 156 (2018) 96–105

Contents lists available at ScienceDirect

Phytochemistry

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

Genetic and functional characterization of Sg-4 glycosyltransferase involved in the formation of sugar chain structure at the C-3 position of soybean saponins

Kyoko Takagi[a,1,2], Ryoichi Yano[a,b,1,3], Saeko Tochigi[a], Yukiko Fujisawa[a,c], Hiroki Tsuchinaga[d], Yuya Takahashi[e], Yoshitake Takada[f], Akito Kaga[a,c], Toyoaki Anai[g], Chigen Tsukamoto[e], Hikaru Seki[d], Toshiya Muranaka[d], Masao Ishimoto[a,c,*]

[a] National Institute of Agrobiological Sciences, Tsukuba, Ibaraki, 305-8602, Japan
[b] Faculty of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Ibaraki, 305-8572, Japan
[c] Institute of Crop Science, NARO, Tsukuba, Ibaraki, 305-8518, Japan
[d] Department of Biotechnology, Graduate School of Engineering, Osaka University, Suita, Osaka, 565-0871, Japan
[e] Graduate School of Agriculture, Iwate University, Morioka, Iwate, 020-8550, Japan
[f] Western Region Agricultural Research Center, NARO, Zentsuji, Kagawa, 765-8508, Japan
[g] Faculty of Agriculture, Saga University, Saga, Saga, 840-8502, Japan

ARTICLE INFO

Keywords:
Glycine max (L.) Merrill
Leguminosae
Genetic analysis
Glycosyltransferase
Triterpenoid saponin

ABSTRACT

Triterpenoid saponins are specialized metabolites, which are abundant in soybean seeds. They have a wide variety of effects on human health and physiology. The composition of sugar chain attached to the aglycone moiety of saponins can be controlled by genetic loci, such as *Sg-1*, *3*, and *4*. Among these, the homozygous recessive *sg-4* impairs the accumulation of saponins that have an arabinose moiety at the second position of the C-3 sugar chain (i.e., saponins Ad and βa) in the hypocotyls. In this study, we found that *sg-4* cultivars are disabled in *Glyma.01G046300* expression in hypocotyls. This gene encodes a putative glycosyltransferase (UGT73P10) and is a homolog of *GmSGT2* (UGT73P2) whose recombinant protein has been previously shown, *in vitro*, to conjugate the second galactose moiety at the C-3 position of soyasapogenol B monoglucuronide (SBMG). The *sg-4* phenotype (absence of saponins Ad and βa in hypocotyls) was restored by introducing the *Glyma.01G046300* genomic DNA fragment that was obtained from the *Sg-4* cultivar 'Ibarakimame 7'. Although *Glyma.01G046300* is expressed in the cotyledons even in the *sg-4* cultivars such as 'Enrei', the induced premature stop codon mutation (W244*) resulted in impaired accumulation of saponin βa in this tissue also in the 'Enrei' genetic background. Furthermore, the recombinant Glyma.01G046300 protein was shown to conjugate the second Ara moiety at the C-3 position of SBMG using UDP-Ara as a sugar donor. These results demonstrate that *Sg-4* is responsible for conjugation of the second Ara moiety at the C-3 position of soybean saponins.

1. Introduction

Seeds of soybean (*Glycine* max (L.) Merr.) (Leguminosae) are major source of essential nutrients in the human diet as well as in the feed of domestic animals around the world. These seeds contain a wide variety of specialized metabolites in addition to beneficial proteins and fats. In particular, triterpenoid saponins are the major specialized metabolites in soybean seeds. Owing to the wide variety of effects of soybean saponins on human health and physiology, the mechanism underlying their biosynthesis has attracted the attention of plant scientists and breeders. Soybean saponins are generally classified into two types based on their chemical structures: group A saponins and 2,3-dihydro-2,5-dihydroxy-6-methyl-4*H*-pyran-4-one (DDMP) saponins (Kudou et al., 1992, 1993; Shiraiwa et al., 1991a; Takada et al., 2013; Tsukamoto et al., 1993; Yano et al., 2017). Group A saponins are bis-desmoside-type saponins that have two sugar chains at the C-3 and C-22

* Corresponding author. Institute of Crop Science, NARO, Tsukuba, Ibaraki, 305-8518, Japan.
E-mail address: ishimoto@affrc.go.jp (M. Ishimoto).
[1] These authors contributed equally to this work.
[2] Present address: Tohoku Agricultural Research Center, NARO, Fukushima, Fukushima 960-2156, Japan.
[3] Present address: Advanced Analysis Center, NARO, Tsukuba, Ibaraki 305-8518, Japan.

https://doi.org/10.1016/j.phytochem.2018.09.002
Received 16 April 2018; Received in revised form 28 August 2018; Accepted 11 September 2018

hydroxyl groups of the aglycone moiety, soyasapogenol A (3β, 21β, 22β, 24-tetrahydroxyolean-12-ene; SA). The DDMP saponins are monodesmoside-type saponins that contain a sugar chain at the C-3 hydroxyl group of the aglycone molecule, in which a DDMP residue forms a hemiacetal linkage with soyasapogenol B (3β, 22β, 24-trihydroxyolean-12-ene; SB) at the C-22 hydroxyl group. The DDMP saponins and their degraded derivatives (group B and group E saponins) have health-promoting functions; they prevent dietary hypercholesterolemia (Fenwick et al., 1991; Murata et al., 2005, 2006), suppress colon cancer cell proliferation (Ellington et al., 2005, 2006), prevent lipid peroxidation, and have hepatoprotective activities resulting from accelerated thyroid hormone secretion (Ishii and Tanizawa, 2006). The acetylated form of group A saponins causes bitter and astringent aftertastes of soy products (Okubo et al., 1992). These properties depend on the chemical structure and concentration of the saponin, and hence, the presence of different saponin components in soybean seeds could confer different health-promoting activities in addition to their effect on the gustatory property of soy food products (Tsukamoto and Yoshiki, 2006).

The composition of sugar chains attached to the soyasapogenol aglycone moiety shows a great variation. The details of the biosynthetic genes involved in the glycosylation processes are not fully understood, as of date. However, previous genetic studies have identified at least seven naturally occurring alleles at three different loci, namely *Sg-1*, *Sg-3*, and *Sg-4* (Kikuchi et al., 1999; Shiraiwa et al., 1990; Takada et al., 2010, 2012; Tsukamoto et al., 1993). Among these, *Sg-1* (*Glyma.07G254600*) has been shown to encode a UGT73 family uridine diphosphate (UDP)-sugar-dependent glycosyltransferase (Sayama et al., 2012). A single amino acid residue at position 138 of the Sg-1 protein sequence determines its substrate specificity; $Sg\text{-}1^{a}$ allele encodes a xylosyltransferase, UGT73F4, whereas $Sg\text{-}1^{b}$ encodes a glucosyltransferase, UGT73F2. The loss-of-function $sg\text{-}1^{0}$ allele results in the absence of the second sugar moiety at the C-22 position of SA, thereby, preventing the generation of acetylated forms of group A saponins. Thus, the $sg\text{-}1^{0}$ allele was used to develop a commercial soybean cultivar, 'Kinusayaka', in which the bitter and astringent group A saponins are absent (Kato et al., 2007). A recessive allele, *sg-3*, was identified as a genetic factor involved in the conjugation of a third glucose (Glc) moiety at the C-3 position of both DDMP and group A saponins (Takada et al., 2012); *sg-3* cultivars fail to accumulate saponins Ab and αg, both of which carry the third Glc moiety at the C-3 position of saponins. Recently, a loss-of-function mutation in *Glyma.10G104700* has been shown to be responsible for the *sg-3* phenotype (Yano et al., 2018). *Sg-3* encodes a glycosyltransferase (UGT91H9) that conjugates UDP-glucose to the third position of the C-3 sugar chain of soybean saponins. *sg*-4 has also been shown to be involved in the conjugation of a second arabinose (Ara) moiety at the C-3 position of both the saponins. Most of the soybean accessions or cultivars widely used in genetic studies, for example, 'Williams 82', 'Enrei', and 'Jack', carry the homozygous recessive *sg-4* allele. These cultivars fail to accumulate saponins Ad and βa, both of which carry the second Ara moiety at the C-3 position, in their hypocotyl (Fig. 1). On the other hand, the Japanese cultivar 'Ibarakimame 7' carries a functional *Sg-4* allele and is, hence, able to accumulate saponins Ad and βa in the hypocotyl. However, in cotyledons, the saponin βa is synthesized in both the *Sg-4* and *sg-4* genotypes, suggesting that the *sg-4* allele affects the accumulation of saponins only in hypocotyls but not in cotyledons (Takada et al., 2012). Herein, the *Sg-4* gene is thought to encode a glycosyltransferase, as does *Sg-3*. In addition to the genetically identified genetic loci, an *in vitro* biochemical study also revealed that GmSGT2 (UGT73P2) and GmSGT3 (UGT91H4) are the enzymes involved in the conjugation of the second galactose (Gal) or the third rhamnose (Rha) moiety at the C-3 position of soyasapogenol B monoglucuronide (SBMG), respectively (Shibuya et al., 2010) (Fig. 1A). A loss-of-function *gmsgt3* mutant was shown to be unable to accumulate saponins with the third Rha at the C-3 sugar chain, and the conjugation of both the third Glc and Rha was disabled in the *sg-3 gmsgt3* double mutant (Yano et al., 2018).

In this study, we identified *Glyma.01G046300* (UGT73P10) as the *Sg-4* gene. The *sg-4* cultivars, namely 'Williams 82', 'Enrei', 'Jack', and 'Suzuyutaka', were found to be disabled in the expression of *Glyma.01G046300* in hypocotyls, whereas it was expressed in 'Ibarakimame 7' (an *Sg-4* genotype). The *sg-4* phenotype (manifested, for example, by the absence of saponins Ad and βa in the seed hypocotyls) was restored by the *Sg-4* genomic fragment obtained from 'Ibarakimame 7' in accordance with the *Glyma.01G046300* gene expression in hypocotyls in the transgenic lines. Furthermore, induced mutations in *Glyma.01G046300* in the *sg-4* cultivar 'Enrei' (W244*) impaired the accumulation of saponin βa in the cotyledons, suggesting that *Glyma.01G046300* plays a role, not only in hypocotyls, but also in cotyledons. The recombinant Glyma.01G046300 protein was also observed to be able to conjugate the second Ara moiety at the C-3 position of saponins by using UDP-Ara as a sugar donor. Finally, an interspecies comparison of UGT73 proteins between soybean and common bean (*Phaseolus vulgaris*) suggested that *Sg*-4 might have specifically evolved in the soybean genome during its evolution.

2. Results and discussion

2.1. Identification of the Sg-4 candidate gene

In a previous study, we mapped the *Sg-4* locus between two SSR markers, GMES0626 and AG77, on Chromosome (Chr.) 1 using recombinant inbred lines (RILs) derived from a cross between 'Ibarakimame 7' (*Sg-4*) and 'Suzuyutaka' (*sg-4*) (Takada et al., 2012). According to the soybean Glyma2.0 genome reference (the *sg-4* cultivar 'Williams 82'), there were 92 genes between the two SSR markers (*Glyma.01G041400* to *Glyma.01G050600*; Fig. 2A). Because *Sg*-4 has been expected to encode an UDP-glycosyltransferase that catalyzes the arabinosylation at the C-3 position of soyasapogenols (Takada et al., 2012; Tsukamoto et al., 1993), we searched for the putative UDP-glycosyltransferase (UGT) gene(s) among these genes according to the Glyma2.0 annotation data (Wm82.a2.v1 gene set). One gene, *Glyma.01G046300*, was found to encode a UGT-like protein (Fig. 2A). The deduced amino acid sequence of *Glyma.01G046300* showed significant homology to that of *GmSGT2* (*Glyma.11G053400*) (BLASTp, 67% sequence similarity, E-value = 0.0) (Fig. 2C). Because GmSGT2 has been previously shown to have a UDP-galactose glycosyltransferase activity on SBMG in an *in vitro* assay (Shibuya et al., 2010), we assumed that Glyma.01G046300 would also have a similar UGT activity on SMBG or other soyasaponin-related compound(s).

We subsequently conducted a Southern blot analysis using a probe amplified from *Glyma.01G046300* to investigate the changes in the genomic structure between the 'Ibarakimame 7' (*Sg-4*) and *sg-4* cultivars ('Williams 82' and 'Suzuyutaka'). The result suggested that the genomic DNA sequences are different between 'Ibarakimame 7' and the two *sg-4* cultivars because 'Ibarakimame 7' carried much shorter genomic DNA fragment relative to 'Williams 82' and 'Suzuyutaka' (Fig. 2B). On the other hand, only one amino acid substitution was present in the deduced protein sequence of *Glyma.01G046300* between 'Ibarakimame 7' (*Sg-4*) and 'Williams 82' (*sg-4*) (P492A, Fig. 2C). Because this substitution occurred at the second codon from behind and also outside of the conserved plant secondary product glycosyltransferase (PSPG) motif that plays an important role in UDP-sugar donor specificity (Gachon et al., 2005; Kubo et al., 2004; Masada et al., 2007), it was unlikely to associate with *sg-4* phenotype at least in 'Williams 82'. We conducted an RT-PCR analysis of both the cotyledon and hypocotyl samples from developing seeds to analyze the expression of *Glyma.01G046300*. When the expression of *Glyma.01G046300* was compared between 'Ibarakimame 7' (*Sg-4*) and four *sg-4* cultivars ('Suzuyutaka', 'Jack', 'Enrei', and 'Williams 82'), 'Ibarakimame 7' showed a different gene expression pattern relative to the four *sg-4* cultivars; significant levels of *Glyma.01G046300* expression were