Related Arabidopsis serine carboxypeptidase-like sinapoylglucose acyltransferases display distinct but overlapping substrate specificities¹

Christopher M. Fraser, Michael G. Thompson¹, Amber M. Shirley², John Ralph³, Jessica A. Schoenherr, Taksina Sinlapadech, Mark C. Hall, and Clint Chapple*

Department of Biochemistry, Purdue University, West Lafayette, IN 47907

³US Dairy Forage Research Center, USDA-Agricultural Research Service, 1925 Linden Drive West, Madison, WI 53706, USA.
This is journal paper number XXXXX from the Purdue University Agricultural Experiment Station. This material is based upon work supported by the National Science Foundation under Grant No. 0091419 to C.C., and a graduate fellowship from Purdue University to C.M.F.

JR gratefully acknowledges partial funding through the DOE Energy Biosciences program (#DE-AI02-00ER15067). NMR experiments on the Bruker DMX-500 cryoprobe system were carried out at the National Magnetic Resonance Facility at Madison with support from the NIH Biomedical Technology Program (RR02301) and additional equipment funding from the University of Wisconsin, NSF Academic Infrastructure Program (BIR-9214394), NIH Shared Instrumentation Program (RR02781, RR08438), NSF Biological Instrumentation Program (DMB-8415048), and the U.S. Department of Agriculture. A portion of the MS analysis was conducted at the Purdue University Metabolic Profiling Facility, supported by the National Science Foundation under grant DBI-0421102.

The authors would like to thank the Salk Institute Genomic Analysis Laboratory (SIGnAL) for providing the sequence-indexed Arabidopsis T-DNA insertion mutants and the Arabidopsis Biological Resource Center (ABRC) for the distribution these materials.

current address, Department of Chemistry, Purdue University, West Lafayette, IN 47907
current address, BASF Plant Science L.L.C., 26 Davis Drive, Research Triangle Park, NC 27709, USA

Corresponding author; e-mail chapple@purdue.edu; fax 765-494-7897.

Abbreviations:

SAT, sinapoylglucose: anthocyanin sinapoyltransferase; SGT, sinapic acid UDPG:sinapoyltransferase; SMT, sinapoylglucose:malate sinapoyltransferase; SCT, sinapoylglucose:choline sinapoyltransferase; SCE, sinapoylcholinesterase; sng, sinapoylglucose accumulator; SST, sinapoylglucose: sinapoylglucose sinapoyltransferase
ABSTRACT

The Arabidopsis genome encodes fifty-one proteins annotated as serine carboxypeptidase-like (SCPL) enzymes. Nineteen of these SCPL proteins are highly similar to one another, and represent a clade that appears to be unique to plants. Two of the most divergent proteins within this group have been characterized to date, sinapoylglucose: malate sinapoyltransferase (SMT) and sinapoylglucose:choline sinapoyltransferase (SCT). The fact that two of the least related proteins within this clade are acyltransferases rather than true serine carboxypeptidases suggests that some or all of the remaining members of this group may have similar activities. The gene that encodes SMT (SNG1: At2g22990) is one of five SCPL genes arranged in a cluster on chromosome 2. In this study, an analysis of deletion mutant lines lacking one or more genes in this SCPL gene cluster reveals that three of these genes also encode sinapoylglucose-dependent acyltransferases. At2g23000 encodes sinapoylglucose: anthocyanin acyltransferase (SAT), an enzyme that is required for the synthesis of the sinapoylated anthocyanins in Arabidopsis. At2g23010 encodes an enzyme capable of synthesizing 1,2-disinapoylgucose from two molecules of sinapoylgucose, an activity shared by SNG1 and At2g22980. Sequence analysis of these SCPL proteins reveals pair-wise percent identities that range from 71% to 78%, suggesting that their differing specificities for acyl acceptor substrates are due to changes in a relatively small subset of amino acids. The study of these SCPL proteins provides an opportunity to examine enzyme structure-function relationships and may shed light on the role of
evolution of hydroxycinnamate ester metabolism and the SCPL gene family in
Arabidopsis and other flowering plants.

INTRODUCTION

Plant secondary metabolites account for a great amount of the
biochemical diversity that exists in Nature. These compounds have been
estimated to number between 100,000 to 200,000, with the majority of them yet
to be studied in detail (Wink, 1988). As a class of molecules, plant secondary
metabolites include compounds that possess many important properties: they
function as antifeedents, phytoanticipins and phytoalexins, signaling molecules,
UV-protectants, and in a host of other physiologically important roles (Li et al.,
1993; Wajant et al., 1994; Kuć, 1995; Kliebenstein et al., 2005; Taylor and
Grotewold, 2005). Aside from their functions within the plant kingdom, they have
also proven to be invaluable to humanity, playing long-standing roles in medicine
and agriculture. Secondary metabolites are therefore not only crucial to our
understanding of plant physiology, but are of great importance to our own health
and survival as well.

A common metabolic step in the production of numerous secondary
metabolites involves the attachment of an acyl group via an activated donor
molecule (Croteau and Hooper, 1978; Dahlbender and Strack, 1986; Villegas
and Kojima, 1986; Villegas et al., 1987; Kesy and Bandurski, 1990; Suzuki et al.,
1994; Lenz and Zenk, 1995; Rabot et al., 1995; Yang et al., 1997; Dudareva et
al., 1998; Fujiwara et al., 1998; Li et al., 1999; Walker et al., 1999; Lehfeldt et al.,
2000; Shirley et al. 2001; Fröhlich et al., 2002; Nakayama et al., 2003). Although
it is well-documented that CoA thioesters often provide the energy necessary for
transacylation reactions in secondary metabolism (St Pierre and De Luca, 2000),
it is now clear that 1-O-glucose esters can serve as activated donor molecules as
well due to their high free energy of hydrolysis (Dahlbender and Strack 1984;
Mock and Strack, 1993). The end-products of these glucose ester-dependent
reactions include many important compounds such as indoleacetyl myo-inositol,
oak gallotannins, isobutyryl glucose esters, certain acylated anthocyanins and
the phenylpropanoid-derived sinapate esters found in Arabidopsis and other
members of the Brassicaceae (Fig. 1) (Kesy and Bandurski, 1990; Gläβgen and
Seitz, 1992; Li et al., 1999; Lehfeldt et al., 2000; Shirley et al., 2001; Fröhlich et
al., 2002).

The enzymes responsible for the final step in the synthesis of isobutyryl
glucose esters in S. berthaultii and L. pennellii, as well as sinapoylcholine and
sinapoylmalate in Arabidopsis (sinapoylglucose:choline sinapoyltransferase
(SCT) and sinapoylglucose:malate sinapoyltransferase (SMT), respectively) are
serine carboxypeptidase-like (SCPL) proteins (Lehfeldt et al., 2000; Li and
Steffens, 2000; Shirley et al., 2001). As a class of enzymes, SCPL proteins bear
the trademark Ser-Asp-His catalytic triad along with other sequence features that
are characteristic of known serine carboxypeptidases (SCPs) such as
carboxypeptidase-Y (CPY) from yeast. In this paper we report the
characterization of three additional Arabidopsis SCPL proteins that are
sinapoylglucose-dependent sinapoyltransferases. The function of these proteins
were identified through the phenotypic characterization of fast neutron-induced
sinapoylglucose accumulator1 (sng1) deletion alleles in which the deletions affect
not only the SNG1 gene, but also one or more of the SCPL genes that flank it.
We report here that At2g23000 encodes a protein that catalyzes the
sinapoylglucose-dependent sinapoylation of anthocyanins in Arabidopsis,
demonstrating that this activity is not dependent upon a BAHD acyltransferase as
has been suggested previously (Luo et al., 2007). Furthermore, we show that the
SCPL protein encoded by At2g23010 catalyzes the disproportionation of two
molecules of sinapoylgucose to generate 1,2-disinapoylgucose (Fig. 1) and an
additional, as yet unidentified compound. We have therefore designated the
SCPL proteins encoded by At2g23000 and At2g23010 as sinapoylglucose:
anthocyanin sinapoyltransferase (SAT) and sinapoylglucose: sinapoylglucose
sinapoyltransferase (SST), respectively. Finally, we show that both SMT and the
protein encoded by At2g22980 are capable of catalyzing the formation of 1,2-
disinapoylgucose, although this activity is clearly not the primary function of
SMT.

Like SCT and SMT, SST, and SAT belong to a family of 51 SCPL proteins
encoded by the Arabidopsis genome, and are members of a clade that includes
15 other closely-related SCPL proteins (Fraser et al., 2005). The fact that SMT,
SST, SAT and the At2g22980 protein are between 71% and 78% identical to one
another (Fraser et al., 2005) suggests that their analysis may permit the
elucidation of structure-function relationships within this subclass of SCPL
proteins. Thus, the SCPL sinapoylglucose acyltransferases (SGAs) represent
additional members of an emerging class of enzymes that catalyze acyltransferase reactions in plant secondary metabolism.

RESULTS

sng1-5 and sng1-6 harbor deletions of an SCPL gene cluster on chromosome 2

The SNG1 gene is one of five SCPL genes that are arranged in tandem on chromosome 2 (Fig. 2). The proteins encoded by these SCPL genes are highly similar, with any two of them being between 71% and 78% identical (Fraser et al., 2005). In order to identify mutations that affect the genes surrounding SNG1, a population of fast neutron-mutagenized plants was screened using the UV-phenotype characteristic of the sng1 mutant (Lehfeldt et al., 2000), and their biochemical sng1 phenotypes were confirmed by HPLC and genetic complementation tests. The genomic regions deleted from each mutant were initially estimated via Southern analysis (Lehfeldt et al., 2000). Additional PCR analysis of the SCPL gene cluster using both gene-specific and intergenic primers indicated that SNG1 (At2g22990) and the gene immediately downstream (At2g23000) were deleted in the sng1-5 mutant. Sequencing of genomic PCR products revealed that the sng1-6 mutant is missing a region of genomic DNA spanning At2g22980 through At2g23010 and four additional non-SCPL genes downstream, none of which encodes an SCPL protein (Fig. 2). In addition, a 170
base pair fragment translocated from gene At4g14980 on chromosome 4 (Blastn E value = 1e -72) is inserted between the 5' and 3' sng1-6 deletion borders.

**Etiolated sng1-6 seedlings lack a major sinapate ester found in etiolated wild-type, sng1-1, and sng1-5 seedlings**

The high sequence similarity shared by SMT and the SCPL proteins encoded by At2g22980, At2g23000, and At2g23010, suggested that one or more of the latter might also be sinapoylglucose-dependent sinapoyltransferases. To test this hypothesis, extracts from a variety of tissues from the sng1-5 and sng1-6 mutants were analyzed via HPLC to determine if any putative sinapate esters present in wild-type and sng1-1 plants were absent from one or both of the mutants. In these experiments, two compounds not previously identified in Arabidopsis with absorption spectra characteristic of sinapate esters were found to be present in wild-type, sng1-1, and sng1-5 seedling extracts, but were absent from sng1-6 extracts (Fig. 3A). Given that At2g22980 and At2g23010 are the only two SCPL genes uniquely deleted from the sng1-6 mutant relative to the sng1-5 mutant, the absence of these compounds (hereafter referred to as compounds 1 and 2) in the sng1-6 seedlings suggested that either At2g22980 or At2g23010 or both are required for their synthesis. Subsequent complementation analyses (see below) confirmed that this is indeed the case. Preliminary liquid chromatography/mass spectrometry (LC/MS) analysis suggested that both compounds were disinapoylated monosaccharides (data not shown), suggesting
that one of these two compounds might be 1,2-disinapoylglucose, which has been shown to accumulate in the cotyledons of radish seedlings (Strack et al., 1984).

**NMR analysis identifies compound 2 as 1,2-disinapoylglucose**

NMR spectra of compound 2 not only established its identity as 1,2-disinapoylglucose but provided the full NMR data and its complete assignment for the first time. The key experiment was the long-range $^{13}\text{C} - ^{1}\text{H}$ (HMBC) experiment, which delineated which sinapoyl unit was attached to the 1- vs 2-position (Fig. 4B). Then, with the coupling networks from each unit revealed in the normal way by COSY and HMBC correlations, all of the protons from each sinapoyl unit (S1 and S2) could be unambiguously assigned, with the possible exception of the methoxyl protons. Short-range 13C–1H (HSQC) correlations unambiguously established the assignments for the protonated carbons (Fig. 4A) and HMBC spectra identified all of the non-protonated carbons. As seen from Fig. 4A, some of the protons are poorly resolved in the 1D proton spectrum, but the proton data are readily extracted from the HSQC spectrum (as well as by COSY, data not shown). The proton and carbon data presented in Table 1 (in two solvents, CD$_3$OD, and 9:1 acetone-d$_6$:D$_2$O) are therefore unambiguously authenticated, again, with the possible exception of the methoxyl protons. Data in Table 1 include previously obtained data (Strack et al. 1984), for comparison. NMR data have not yet been obtained to ascertain the structure of compound 1 due to its very low abundance in extracts.
**At2g23010 complements the sng1-6 phenotype**

To test the hypothesis that At2g23010 is required for the synthesis of compound 1 and 1,2-disinapoylglucose, the sng1-6 mutant was transformed with pCC579, a vector containing the genomic upstream, downstream, and coding regions of the gene. HPLC analysis of extracts from etiolated, 6-day-old sng1-6/pCC579 seedlings showed that both compound 1 and 1,2-disinapoylglucose are present at wild-type levels in the transformed mutant seedlings, indicating that At2g23010 complements this aspect of the sng1-6 phenotype (Fig. 3A), and that this gene thus encodes SST.

To determine if the protein encoded by At2g23010 is uniquely responsible for the synthesis of 1,2-disinapoylglucose in etiolated seedlings, we obtained a SALK Institute T-DNA line (SALK_133207) harboring an insertion in At2g23010. Homozygosity for the T-DNA insertion was initially determined by screening for kanamycin resistance and by PCR genotyping, and Southern analysis was then used to confirm these results (data not shown). Although compound 1 is absent from extracts of etiolated SALK_133207 seedlings, 1,2-disinapoylglucose is still accumulated in the mutants seedlings (Fig. 3B), albeit at lower levels than in the wild type. These data suggested that one or more of At2g22980, SNG1 or At2g23000 are capable of synthesizing 1,2-disinapoylglucose.

**At2g22980 and SNG1 may contribute to 1,2-disinapoylglucose accumulation in vivo**
To test the hypothesis that one or more of the SCPL genes clustered near At2g23010 also have SST activity, two additional *sng1-6* transgenic lines were generated: one harboring SNG1 (pCC398), and one harboring At2g22980 (pCC881) under the control of their native regulatory elements. Extracts from 6-day-old etiolated seedlings of each transgenic line contained 1,2-disinapoylglucose but not compound 1 (Fig. 3B), thus exhibiting a phenotype similar to that of the SALK At2g23010 knock-out line. These data show that in addition to the At2g23010 protein, both of these enzymes have the capacity to synthesize 1,2-disinapoylglucose.

In addition to the observed redundancy with respect to the genes missing from the *sng1-6* mutants, results from enzyme assays suggest that another Arabidopsis protein, likely an SCPL enzyme, exhibits SST activity *in vitro*. We carried out enzyme assays using crude protein extracts from 6-day-old, etiolated wild-type and *sng1-6* seedlings, and found roughly equivalent SST activity for both (data not shown). Given that at least five other Clade 1 SCPL genes are expressed in seedlings (Fraser et al., 2005) and that the proteins encoded by these genes are similar to SST, it is reasonable to expect that one of these SCPL proteins might be able to catalyze the disproportionation reaction leading to the production of 1,2-disinapoylglucose *in vitro*. On the other hand, the lack of 1,2-disinapoylglucose in the *sng1-6* mutant suggests that the corresponding enzyme(s) may not be expressed in cells in which sinapoylglucose is available as a substrate.
1,2-Disinapoylglucose and compound 1 accumulate in *sng1-1* and *sng1-5* leaves but not in wild-type or *sng1-6* leaves

Although At2g23010 is expressed in wild-type leaves (Fraser et al., 2005), HPLC analysis shows that 1,2-disinapoylglucose is not present in wild-type leaf extracts (Fig. 5), suggesting that the expression of SST may be rendered moot by higher levels of SMT activity. In contrast, 1,2-disinapoylglucose is present in the leaves of the *sng1-1*, and *sng1-5* mutants, suggesting that the cryptic levels of SST expression can be revealed when sinapoylglucose levels are not depleted by SMT as occurs in a wild-type background. Consistent with the hypothesis that At2g23010 encodes SST, 1,2-disinapoylglucose is not found in *sng1-6* leaves (Fig. 5).

In the course of investigating the *bright trichomes 1* (*brt1*) mutant, which is defective in a major sinapic acid: UDPG glucosyltransferase (Sinlapadech et al., 2007), we generated a *brt1-1/sng1-1* double mutant. HPLC analysis of the double mutant showed that total sinapate ester content in the *brt1-1/sng1-1* leaf extracts was not substantially different from that of the *brt1-1* mutant; however, 1,2-disinapoylglucose levels were dramatically reduced compared to *sng1-1* (Fig. 6). Taken together, these data suggest that SMT and SST are probably expressed in the same cell types where they compete with one another for substrate. Furthermore, they suggest that BRT1 is the major SGT expressed in cells that synthesize 1,2-disinapoylglucose, and conversely, that 1,2-disinapoylglucose is synthesized in only a subset of cells that accumulate sinapoylglucose.
The *sng1-5* and *sng1-6* mutants lack the sinapoylated form of the major anthocyanin in Arabidopsis

The most abundant anthocyanin in Arabidopsis is a sinapoylated cyanidin derivative, 3-O-[6-O-(4-O-β-D-glucopyranosyl-β-coumaroyl)-2-O-(2-O-sinapoyl-β-D-xylopyranosyl)-β-D-glucopyranosyl]-5-O-(6-O-malonyl-β-D-glucopyranosyl)cyanidin (Fig. 1; Bloor and Abrahams, 2002; anthocyanin A11 according to the nomenclature of Tohge et al., 2005). Interestingly, visual inspection of 4-day-old light-grown seedlings revealed that whereas both wild-type and *sng1-1* seedlings frequently show pigmentation in the upper portion of the hypocotyl and cotyledons, *sng1-5* and *sng1-6* plants never exhibit this coloration (data not shown). RT-PCR expression analysis was also consistent with the hypothesis that At2g23000 might have a role in anthocyanin biosynthesis since the gene is expressed in senescing leaves (Fraser et al., 2005), a tissue that is frequently observed to accumulate anthocyanin pigments. The At2g23000 gene is also upregulated in the *pap1-D* mutant, which accumulates anthocyanins to abnormally high levels (Tohge et al., 2005). To induce anthocyanin accumulation in a convenient and reproducible way, we stressed MS-grown seedlings by exposure to 10% sucrose for five days, which produced a distinct anthocyanin-less phenotype in the *sng1-5* and *sng1-6* mutants relative to wild type and *sng1-1*. HPLC analysis of acidic methanol extracts of these sucrose-stressed wild-type and *sng1-1* seedlings revealed the presence of a major anthocyanin that is absent from both *sng1-5* and *sng1-6*. 
seedlings (Fig. 7). Analysis of this compound by MALDI-TOF yielded data that identified it as A11, the major anthocyanin in Arabidopsis (Bloor and Abrahams, 2002; Tohge et al., 2005). Furthermore, although wild-type Arabidopsis accumulates a number of sinapoylated anthocyanins (Bloor and Abrahams, 2002; Tohge et al., 2005), none of these compounds is present in the sng1-5 mutant, suggesting that sng1-5 lacks the enzyme required for the production of sinapoylated anthocyanins. Since the only difference between the sng1-1 and sng1-5 mutants is the absence of At2g23000 in sng1-5, these data suggest that At2g23000 encodes sinapoylglucose:anthocyanin sinapoyltransferase (SAT).

At2g23000 is uniquely responsible for anthocyanin sinapoylation in vivo

The data above do not exclude the possibility that SMT is also an SAT, and that SMT and At2g23000 are redundant in function, at least with respect to anthocyanin sinapoylation. Since, a T-DNA insertional mutant for At2g23000 is not currently available, the sng1-5 mutant was transformed with pCC398, the aforementioned vector harboring the SNG1 gene, to determine the effect of a loss of At2g23000 function alone. Twelve independent transformed lines from the T2 generation were obtained and screened via HPLC. Levels of sinapoylmalate for each of transformed lines varied, but in no case were sinapoylated anthocyanins detected (Fig. 8). The presence of sinapoylmalate in these plants indicate that SMT is expressed and catalytically active, and the lack of sinapoylated anthocyanin demonstrates that the enzyme is incapable of acting as a redundant SAT in vivo.
DISCUSSION

SCPL proteins represent a class of proteins whose diversity of function has become apparent only recently. SCPL proteins have been designated as serine carboxypeptidase-like solely on the basis of their overall sequence similarity with respect to known SCPs but the enzymatic function of SCPL proteins implied by their annotation is entirely assumed. While the annotation is in some cases accurate, as with the Arabidopsis SCPL protein BRS1 (Zhou and Li, 2005), it is now clear that as a class, SCPL proteins are diverse in both the reactions they catalyze and the functions they serve. Indeed, the SCPL hydroxynitrile lyase from Sorghum bicolor (SbHNL) (Wajant et al., 1994), the SCPL isobutyrylglucose acyltransferases from L. pennellii (Li and Steffens, 2000), and the SCPL-SGAs from Arabidopsis and other members of the Brassicaceae all illustrate that SCPL proteins exhibit a range of catalytic activities (Lehfeldt et al., 2000, Shirley et al., 2001, Milkowski et al., 2004).

The Arabidopsis SCPL-SGA gene cluster exemplifies a mechanism underpinning the diversification of plant secondary metabolism

Considering that the genes encoding SMT, SAT, SST and the At2g22980 protein are arranged in a cluster, their divergence in substrate and product specificity serves as an excellent example of how gene duplication can contribute to metabolic diversity, and illustrates a means by which secondary metabolism in plants can evolve and diversify. Genes of secondary metabolism arranged in
tandem often encode proteins that are similar in metabolic roles, substrate specificity, or catalytic activity (Frey et al., 1997; Hamberger and Hahlbrock, 2004; Bowles et al., 2005) but often exhibit distinct biochemical and biological functions. For example, \textit{Bx2} through \textit{Bx5}, maize genes involved in the biosynthesis of the natural pesticide 2,4-dihydroxy-7-methoxy-1,4-benzoxazin-3-one (DIMBOA) are clustered on the short arm of chromosome 4. These genes encode related cytochrome P450-dependent monooxygenases, each of which in turn metabolizes sequential DIMBOA biosynthetic intermediates (Frey et al., 1997). Similarly, a number of Arabidopsis UDP-glucosyltransferases (UGTs) are encoded by tandemly-arranged genes retain the same catalytic activity and donor molecule specificity for UDP-glucose but differ in their acceptor substrate specificity (Bowles et al., 2005). For example, UGT73C6 has been identified as a UDP-glucose:flavonol-3-O-glycoside-7-O-glucosyltransferase (Jones et al., 2003); whereas, UGT73C5 has been identified as a deoxynivalenol glucosyltransferase (Poppenberger et al., 2003). The genes encoding these two enzymes are neighbors in a six-member gene cluster at the bottom of chromosome II. In contrast, UGT84A1 and UGT84A3, which are encoded by members of a gene cluster on chromosome IV, exhibit activity toward a broad range of hydroxycinnamic acids, with UGT84A1 and UGT84A3 showing the highest catalytic efficiencies for caffeic acid and ferulic acid, respectively (Lim et al., 2001). The SCPL-SGA gene cluster that includes SMT reflects aspects of both of these latter examples in that SST, SMT and the protein encoded by
At2g22980 are all sinapoylglucose-dependent SGAs capable of producing both 1,2-disinapoylglucose and distinct products as well.

The evolution of the SCPL-SGAs is likely to have taken place by an initial shift in catalytic activity from hydrolytic cleavage exhibited by SCPs to acyltransferase activity, followed by divergence of substrate specificity. The initial shift in catalytic activity may also represent the point of divergence in the function of the ancestral SCP from involvement in primary cellular processes (e.g. protein processing and degradation), providing the opportunity for members of the ancestral gene family to assume a role in plant secondary metabolism. Although a distinct function for At2g22980 has not yet been identified, and the principle role for SMT is clearly in sinapoylmalate biosynthesis, the partial redundancy of these proteins with SST indicates that the evolution of these genes has not yet resulted in enzymes with completely distinct acyl acceptor specificities. In contrast, the ability of SMT, SAT, and SST to function uniquely in the biosynthesis of sinapoylmalate, sinapoylated anthocyanins, and compound 1, respectively, appears to be shared by no other members of the protein family encoded by this gene cluster. In summary, this SCPL-SGA gene cluster appears to provide a glimpse of the gene evolution underlying plant secondary metabolism.

It is interesting to note that the SCPL-SGA gene cluster on chromosome II is one of seven tandem clusters found within the Arabidopsis SCPL gene family, with the three largest clusters all encoding Clade IA SCPL proteins (Fraser et al., 2005). In fact, only two of nineteen SCPL Clade IA genes are found outside of
these tandem gene clusters. In light of the results presented in this paper, and the fact that the Clade IA proteins are highly similar to one another, it is reasonable to expect that these other SCPL proteins not only function as SCPL-SGAs, but that they may overlap in acyl-donor and/or acceptor specificity as well. This is especially likely for the SCPL proteins encoded by the Clade IA SCPL gene cluster located on lower arm of chromosome I, given that they encode the most similar group of all SCPL proteins, with pair-wise percent identities between 79 and 86% (Fraser et al., 2005). The presence of multiple gene clusters encoding highly similar SCPL proteins is also consistent with the hypothesis that Clade IA SCPL proteins have evolved relatively recently and the that they are, as a class, associated with plant secondary metabolism.

**Sinapoylation by SCPL-SGAs contributes to the diversity of plant secondary metabolites**

Arabidopsis and other members of the *Brassicaceae* synthesize a wide array of sinapate esters in addition to those discussed here (Baumert et al., 2005). While the physiological roles of many of these compounds have yet to be determined, as with 1,2-disinapoylglucose (compound 2) and compound 1, their abundance and diversity make sinapate esters an interesting class of plant secondary metabolites, and it is likely many of these compounds are produced via SCPL-SGAs (Lehfeldt et al., 2000, Shirley et al., 2001, Baumert et al., 2005; Fraser et al., 2005). Anthocyanins, on the other hand, are wide-spread throughout the plant kingdom, and play significant roles in pollination, seed
dispersal, and plant stress responses (Winkel-Shirley, 2001; Winkel-Shirley, 2002), so gaining an understanding of the steps involved in their biosynthesis, including acylation, is important to understanding plant biochemistry on a more general level. Anthocyanins vary widely in their patterns of acylation, with activated hydroxycinnamates commonly serving as acyl donors. It has been suggested that acyl groups may contribute to the stabilization of anthocyanins by allowing for molecular stacking, thereby preventing nucleophilic attack by water and preserving color (Brouillard, 1981; Brouillard, 1983; Hopp and Seitz, 1987). Acylation also affects the absorbance properties of anthocyanins, with different patterns of acylation resulting in observable differences in color (Honda et al., 2005), and it has been suggested that pH-dependent conformational changes brought about by the presence of acyl groups may be involved in localization of anthocyanins to the vacuole (Matern et al., 1986; Hopp and Seitz, 1987). Understanding the role that SCPL-SGAs may play in the acylation of anthocyanins may therefore provide important insights not only into plant secondary metabolism, but plant survival and adaptability as well.

Sinapoylated anthocyanins have been found in many different members of the Brassicaceae, including cabbage (Hrazdina et al., 1977; Wu and Prior, 2005), *Matthiola incana* (Saito et al., 1995; Saito et al., 1996), *Sinapis alba* (Takeda et al., 1988) and *Orychophragonus violaceus* (Honda et al., 2005). Sinapoylated anthocyanins are also found in some orchids (Tatsuzawa et al., 2004), and the major anthocyanin in wild carrot is sinapoylated (Harborne et al., 1983) where sinapoylglucose has been shown to serve as the activated acyl donor molecule.
(Gläβgen and Seitz, 1992). Based upon our findings, anthocyanin sinapoylation in some, or possibly all of these plants may be catalyzed by SCPL-SGAs and the acylation of anthocyanins with other hydroxycinnamates via SCPL-SGAs may serve as alternative catalysts to the more common BAHD acyltransferases (St Pierre and De Luca, 2000; Luo et al., 2007).

The metabolic route to sinapoylation still remains to be determined. While it is possible that SAT exhibits substrate specificity towards a single anthocyanin such as 3-O-(6-O-p-coumaroyl-2-O-β-D-xylopyranosyl-β-D-glucopyranosyl)5-O-(6-O-malonyl-β-D-glucopyranosyl)cyanidin, designated A5 in Tohge et al., 2005), an alternative is that the enzyme may be able to use multiple substrates as acyl acceptors (Fig. 9). For instance both A5 and A8 may act as acyl acceptors in the biosynthesis of A9 and A11, respectively (Fig. 9). A similar scenario appears to exist for the acyltransferases that add the malonyl and p-coumaroyl subunits to the glycosylated form of the core anthocyanin molecule, resulting in an anthocyanin biosynthetic metabolic grid (Fig. 9) (Luo et al., 2007).

**Substrate specificity of the Arabidopsis SCPL-SGAs may be due to a small subset of amino acid residues**

SST, SAT and the At2g22980 protein represent the latest additions to the class of SCPL-SGAs whose activities have been fully or partially characterized. As such, they not only expand the importance of this class of enzymes with respect to plant secondary metabolism, but they also provide an opportunity to study the relationship between the evolution of sequence and function. For
example, the fact that SMT, SST, SAT, the protein encoded by At2g22980, and even SCT are relatively similar to one another would suggest that the relatively small number of amino acids by which these proteins differ is responsible for the differing substrate specificities exhibited by the Arabidopsis SCPL-SGAs, and that the most important amino acids affecting substrate specificity are likely to be an even smaller subset of residues clustered near the enzymes' active sites.

In conclusion, the SCPL-SGAs represent a growing class of enzymes important to plant secondary metabolism. Our research and the research of others suggests that in the plant kingdom, a significant number SCPL proteins have been recruited to serve as glucose acyltransferases, utilizing 1-O-sinapoylglucose as an activated acyl donor molecule. The results of this paper further suggest that the study of the SCPL-SGA gene cluster provides an opportunity to elucidate the relationship between gene sequence, enzyme function, and the evolution of metabolism.

MATERIALS AND METHODS

Plant material and growth conditions

Arabidopsis plants were grown under a 16 hr light/ 8 hr dark photoperiod in soilless potting mix (Redi-Earth potting mixture, Scotts-Sierra Horticultural Products; Marysville, OH) at 23°C. For seedling plant material to be used in the analysis of sinapate ester accumulation, seeds were surface sterilized for 10 min in a 2:1 mixture of 0.1% Triton X-100 and household bleach. Seeds were rinsed
thoroughly with sterile water and planted on modified Murashige and Skoog medium (Murashige and Skoog, 1962) (ammonia-free medium to which 20.6 mM potassium nitrate was added in place of ammonium nitrate) containing 0.7% agar. For seedling plant material to be used for the isolation of 1,2-disinapoylglucose, non-sterilized seeds were sown on filter paper moistened with water. In order to induce anthocyanin production, seedlings were grown in liquid MS media containing 1% sucrose for 6 days, the media was then removed and replaced with 10% sucrose MS media and the seedlings were grown for an additional 5 days.

**Analysis of sinapate esters and anthocyanins**

Sinapate esters were extracted from fresh plant tissue in 50% (v/v) methanol. Samples were incubated at 65°C for 30 minutes, centrifuged at 12,000g for 2 min, and analyzed by high-performance liquid chromatography (HPLC) on a Microsorb-MV C\textsubscript{18} column (Varian, Lake Forest, CA) using a gradient from 10% acetonitrile in 1.5% acetic acid to 35% acetonitrile in 1.5% acetic acid at a flow rate of 1 mL min\textsuperscript{-1}. Sinapate esters were detected by their UV absorbance at 335 nm. Anthocyanins were extracted from fresh plant tissue in 4:46:50 acetic acid:water:methanol. Samples were incubated at 65°C for 30 minutes, centrifuged at 12,000g for 2 min, and analyzed by HPLC on an end-capped C\textsubscript{18} column (Waters, Milford, MA) using a gradient from 10% acetonitrile in 10% formic acid to 20% acetonitrile in 10% formic acid flow rate of 1 mL min\textsuperscript{-1}. Anthocyanins were detected by their UV absorbance at 525 nm.
fractions corresponding to each peak were collected and analyzed by matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry (MS).

**Isolation and identification of 1,2-disinapoylglucose (compound 2)**

For the identification of the sinapate esters not found in the *sng1-6* mutant, *sng1-1* seedlings were grown on moistened filter paper for seven days in complete darkness and were extracted in 50% methanol. This extract was concentrated *in vacuo*, and analyzed using HPLC as described above. The pooled fractions containing 1,2-disinapoylglucose were again reduced *in vacuo* to 2 mL and analyzed by NMR.

A sub-sample (ca. 1 mg) of the putative 1(\(E\)),2(\(E\))-di-\(O\)-sinapoyl-\(\beta\)-D-glucopyranoside (1,2-disinapoyl glucose) isolated by HPLC was dissolved in either CD\(_3\)OD or 9:1 acetone-d\(_6\):D\(_2\)O and transferred to a 5 mm OD NMR tube. The usual array of 1D (1H, 13C) and 2D (COSY, HSQC, HSQC-TOCSY, HMBC) experiments were run using standard Bruker pulse programs, from which unambiguous assignments were established. The spectra in perdeutero-methanol were acquired on a Bruker DRX-360 using an inverse (proton coils closest to the sample), \(^1\)H/broadband, gradient probe. The spectra in perdeutero-acetone:D\(_2\)O were acquired on a Bruker DMX-500 equipped with an inverse, gradient, cryogenically-cooled probe for enhanced sensitivity. HMBC spectra were acquired with both an 80 and 120 ms long-range coupling delay.
Construction of transgenic lines

To construct a plant transformation vector containing At2g23010 (pCC579) a cosmid carrying the corresponding genomic sequence (pCC301) was digested with *Nsi* I to liberate a 7500 bp fragment that contained the genomic coding sequence for At2g23010 as well as 2200 bp upstream and 1628 bp downstream of the At2g23010 open reading frame. This fragment was subsequently subcloned into the compatible *Pst* I site of the binary vector, pCAMBIA 2300, generating pCC579. To construct the plasmid containing At2g22980 a cosmid carrying the corresponding genomic sequence (pCC305) was digested with *Hind*III and *Xba*I to liberate a 5865bp fragment that contained the genomic coding sequence of At2g22980, 1066bp upstream and 741bp downstream of the At2g22980 open reading frame. The plasmid containing At2g22990 (*SNG1*) was constructed as described elsewhere (Lehfeldt et al., 2001).

Constructs for plant transformation were introduced into *Agrobacterium tumefaciens* C58 pGV3850 (Zambryski et al., 1983) by electroporation. Plant transformation was performed by the floral dip method (Clough and Bent, 1998) into *sng1*-6. Transformed seedlings (T₁) identified by selection on Murashige-Skoog medium containing 50 mg L⁻¹ kanamycin and 200 mg L⁻¹ timentin were transformed to soil.

Matrix-assisted laser desorption ionization (MALDI) mass spectrometry
Samples recovered after HPLC were spotted on a stainless steel MALDI plate with an equal volume of 5 mg/ml α-cyano-4-hydroxycinnamic acid in 50 % acetonitrile/0.1 % trifluoroacetic acid and allowed to air dry. Mass spectra were acquired with an Applied Biosystems 4700 MALDI TOF/TOF instrument operating in positive reflectron mode. The instrument was initially calibrated with a series of peptide standards to a mass accuracy better than 50 ppm.
LITERATURE CITED


Analysis of a chemical plant defense mechanism in grasses. Science 277: 696-699


Gläßgen WE, Seitz HU (1992) Acylation of anthocyanins with hydroxycinnamic acids via 1-O-acylglucosides by protein preparations from cell cultures of Daucus carota L. Planta 186: 582-585


Table 1. $^{13}$C-NMR data for disinapoyl glucose derived from the usual array of 1D and 2D experiments, and comparison (assigned here) with prior data (Strack et al., 1984). Ac, acetone-D$_6$.

<table>
<thead>
<tr>
<th>Carbon</th>
<th>δ MeOD</th>
<th>δ Ac/D$_2$O</th>
<th>δ* MeOD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>94.2</td>
<td>93.2</td>
<td>94.1</td>
</tr>
<tr>
<td>2</td>
<td>74.4</td>
<td>73.5</td>
<td>74.4</td>
</tr>
<tr>
<td>3</td>
<td>76.2</td>
<td>75.2</td>
<td>76.1</td>
</tr>
<tr>
<td>4</td>
<td>71.4</td>
<td>70.8</td>
<td>71.4</td>
</tr>
<tr>
<td>5</td>
<td>79.3</td>
<td>78.5</td>
<td>79.2</td>
</tr>
<tr>
<td>6</td>
<td>62.4</td>
<td>61.8</td>
<td>62.4</td>
</tr>
</tbody>
</table>

S1 sinapoyl unit

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>126.3</td>
<td>125.3</td>
<td>126.4</td>
</tr>
<tr>
<td>2/6</td>
<td>107.3</td>
<td>106.9</td>
<td>107.3</td>
</tr>
<tr>
<td>3/5</td>
<td>149.6</td>
<td>148.5</td>
<td>149.5</td>
</tr>
<tr>
<td>4</td>
<td>140.3</td>
<td>139.7</td>
<td>140.2</td>
</tr>
<tr>
<td>7</td>
<td>149.1</td>
<td>148.3</td>
<td>149.0</td>
</tr>
<tr>
<td>8</td>
<td>114.7</td>
<td>114.2</td>
<td>114.7</td>
</tr>
<tr>
<td>9</td>
<td>167.4</td>
<td>166.1</td>
<td>167.1</td>
</tr>
<tr>
<td>3/5-OMe</td>
<td>56.9</td>
<td>56.5</td>
<td>56.9</td>
</tr>
</tbody>
</table>

S2 sinapoyl unit

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>126.7</td>
<td>125.7</td>
<td>126.6</td>
</tr>
<tr>
<td>2/6</td>
<td>107.1</td>
<td>106.6</td>
<td>107.1</td>
</tr>
<tr>
<td>3/5</td>
<td>149.6</td>
<td>148.7</td>
<td>149.5</td>
</tr>
<tr>
<td>4</td>
<td>139.8</td>
<td>139.3</td>
<td>139.9</td>
</tr>
<tr>
<td>7</td>
<td>148.0</td>
<td>146.9</td>
<td>147.8</td>
</tr>
<tr>
<td>8</td>
<td>115.7</td>
<td>115.2</td>
<td>115.6</td>
</tr>
<tr>
<td>9</td>
<td>168.6</td>
<td>166.9</td>
<td>168.3</td>
</tr>
<tr>
<td>3/5-OMe</td>
<td>56.9</td>
<td>56.5</td>
<td>56.9</td>
</tr>
</tbody>
</table>

* Data from Strack et al., 1984.
Table 2. $^1$H-NMR data for disinapoyl glucose derived from the usual array of 1D and 2D experiments, and comparison (assigned here) with prior data (Strack et al., 1984).

<table>
<thead>
<tr>
<th>Proton</th>
<th>$\delta$ (m, J) MeOD</th>
<th>$\delta$ (m, J) Ac/D$_2$O</th>
<th>$\delta^*$ MeOD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5.80 (d, 8.4)</td>
<td>5.75 (d, 8.4)</td>
<td>5.85</td>
</tr>
<tr>
<td>2</td>
<td>5.07 (dd, 9.5, 8.4)</td>
<td>5.03 (dd, 9.7, 8.4)</td>
<td>5.12</td>
</tr>
<tr>
<td>3</td>
<td>3.73 (m)</td>
<td>3.77 (m)</td>
<td>3.77</td>
</tr>
<tr>
<td>4</td>
<td>3.518 (-)</td>
<td>3.53 (-)</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>3.515 (-)</td>
<td>3.53 (-)</td>
<td>-</td>
</tr>
<tr>
<td>6a</td>
<td>3.76 (-)</td>
<td>3.70 (-)</td>
<td>-</td>
</tr>
<tr>
<td>6b</td>
<td>3.89 (-)</td>
<td>3.84 (-)</td>
<td>3.95</td>
</tr>
</tbody>
</table>

**S1 sinapoyl unit**

<table>
<thead>
<tr>
<th></th>
<th>$\delta$ MeOD</th>
<th>$\delta$ Ac/D$_2$O</th>
<th>$\delta^*$ MeOD</th>
</tr>
</thead>
<tbody>
<tr>
<td>2/6</td>
<td>6.89 (s)</td>
<td>6.96 (s)</td>
<td>6.91</td>
</tr>
<tr>
<td>7</td>
<td>7.64 (d, 16.0)</td>
<td>7.60 (d, 15.9)</td>
<td>7.67</td>
</tr>
<tr>
<td>8</td>
<td>6.32 (d, 16.0)</td>
<td>6.33 (d, 15.9)</td>
<td>6.36</td>
</tr>
<tr>
<td>3/5-OMe</td>
<td>3.86** (s)</td>
<td>3.83** (s)</td>
<td>3.89**</td>
</tr>
</tbody>
</table>

**S2 sinapoyl unit**

<table>
<thead>
<tr>
<th></th>
<th>$\delta$ MeOD</th>
<th>$\delta$ Ac/D$_2$O</th>
<th>$\delta^*$ MeOD</th>
</tr>
</thead>
<tbody>
<tr>
<td>2/6</td>
<td>6.85 (s)</td>
<td>6.91 (s)</td>
<td>6.88</td>
</tr>
<tr>
<td>7</td>
<td>7.63 (d, 16.0)</td>
<td>7.58 (d, 15.9)</td>
<td>7.66</td>
</tr>
<tr>
<td>8</td>
<td>6.41 (d, 16.0)</td>
<td>6.37 (d, 15.9)</td>
<td>6.45</td>
</tr>
<tr>
<td>3/5-OMe</td>
<td>3.84** (s)</td>
<td>3.81** (s)</td>
<td>3.87**</td>
</tr>
</tbody>
</table>

* Data from Strack et al., 1984; We use the central CHD$_2$OD peak as reference $\delta=3.31$ ppm; our data differ by $\sim0.04$ ppm from Strack’s.

** Assignments may be interchanged.
Table 3. Identity and function of characterized SCPL acyltransferases in Arabidopsis. SCPL # designation refers to the nomenclature developed by Fraser et al., (2005).

<table>
<thead>
<tr>
<th>AGI code</th>
<th>SCPL #</th>
<th>Major activity</th>
<th>Minor activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>At2g22980</td>
<td>13</td>
<td>-</td>
<td>SST*</td>
</tr>
<tr>
<td>At2g22990</td>
<td>8</td>
<td>SMT</td>
<td>SST</td>
</tr>
<tr>
<td>At2g23000</td>
<td>10</td>
<td>SAT</td>
<td>-</td>
</tr>
<tr>
<td>At2g23010</td>
<td>9</td>
<td>SST</td>
<td>-</td>
</tr>
<tr>
<td>At5g09640</td>
<td>19</td>
<td>SCT</td>
<td>-</td>
</tr>
</tbody>
</table>

*SST is the only known activity for At2g22980. Considering that future experiments may reveal that its prime role is in the synthesis of a compound other than 1,2-disinapoylglucose, SST is listed as a minor activity for this protein.
FIGURE LEGENDS

Figure 1. The pathway of sinapate ester biosynthesis in Arabidopsis.

Multiple steps are required for the conversion of phenylalanine to sinapic acid via $p$-coumaroyl-CoA. The enzymes required for the synthesis of sinapate esters in developing seeds are UDP-glucose: sinapic acid glucosyltransferase (SGT), and sinapoylglucose: choline sinapoyltransferase (SCT). Upon germination, sinapoylcholinesterase (SCE) catalyzes the hydrolysis of seed reserves of sinapoylcholine, and SGT and sinapoylglucose: malate sinapoyltransferase (SMT) convert the released sinapate into sinapoylmalate. Sinapoylglucose: sinapoylglucose sinapoyltransferase (SST) catalyzes the conversion of two molecules of sinapoylglucose to 1,2-disinapoylglucose and compound 1, a biosynthetic route that is predominant in etiolated seedlings. Dashed arrows indicate one possible route to the biosynthesis of compound 1, a reaction dependent upon SST. The action of sinapoylglucose:anthocyanin sinapoyltransferase (SAT) is required for the sinapoylation of Arabidopsis anthocyanins. Double arrows proceeding from $p$-coumaroyl-CoA indicate the biosynthesis of the cyanidin derivative via flavonoid metabolism and the subsequent acylation of the anthocyanin, presumably by a $p$-coumaroyl-CoA dependent acyltransferase. This biosynthetic pathway yields one possible substrate for SAT, $3'-O-(6'-O-p$-coumaroyl-$2'$-$O$-$\beta$-$D$-xylopyranosyl$-\beta$-$D$-glucopyranosyl)$-5'$-$O-(6'-O$-malonyl-$\beta$-$D$-glucopyranosyl)$cyanidin$ (A5;
nomenclature from Tohge et al. 2005), sinapoylation of which would result in the production of 3-O-[6-O-(4-O-β-D-glucopyranosyl-p-coumaroyl)-2-O-(2-O-sinapoyl-β-D-xylopyranosyl)-β-D-glucopyranosyl]-5-O-(6-O-malonyl-β-D-glucopyranosyl)cyaniadin (A11)

Figure 2. The sng1 genotypes and the region surrounding the SNG1 locus.

The SNG1 gene encodes SMT, and is one of five SCPL genes (white boxes) in a tandem cluster on chromosome 2. The deletions in the sng1-5 and sng1-6 mutants are designated by double headed arrows, and include a number of non-SCPL genes (grey boxes) in sng1-6. Of the two SCPL genes uniquely deleted in sng1-6, At2g22980 and At2g23010, the latter encodes SST and its deletion is responsible for the lack of compound 1 accumulation in sng1-6. The combined absence of At2g22980, SMT, and At2g23010 explains the absence of 1,2-disinapoylglucose in the mutant. The deletion of At2g23000 in sng1-5 and sng1-6 results in the inability of these mutants to accumulate sinapoylated anthocyanins. Triangles indicate the sites of the EMS mutation in sng1-1 and the site of the T-DNA insertion in the SALK_133207 mutant.

Figure 3. Analysis of the sinapate esters in the sng1, SALK_133207, and transformed sng1-6 mutants.
Extracts were prepared from 7-day old, etiolated dark-grown seedlings and analyzed by HPLC with UV detection at 335 nm. A, extracts from Columbia wild-type, sng1-1, sng1-5, and sng1-6 mutants as well as the sng1-6 mutant transformed with the At2g23010 genomic construct (pCC579). B, extracts from the T-DNA insertional mutant of At2g23010 (SALK_133207), and sng1-6 transformed with either the At2g22980 genomic construct (pCC881) or the At2g22990/SMT construct (pCC398). The sng1 mutants are defective in SMT and accumulate sinapoylglucose (SG) in their leaves instead of accumulating the sinapoylmalate (SM) that is found in wild-type plants. Compounds 1 (a disinapoylated monosaccharide) and 1,2-disinapoylglucose do not accumulate in the sng1-6 mutant, both compounds accumulate in the sng1-6/pCC579 mutant, but only 1,2-disinapoylglucose accumulates in SALK_133207, sng1-6/pCC881 and sng1-6/pCC398.

**Figure 4.** NMR analysis of 1,2-disinapoylglucose isolated from Arabidopsis.

Spectra from isolated 1,2-disinapoylglucose in 9:1 acetone-d₆:D₂O. A, HSQC spectrum (with the 1D proton spectrum as the “projection” on top) showing all fully-resolved short-range (1-bond) $^{13}$C→$^1$H correlations. B, Small section of the HMBC (120 ms long-range coupling delay) highlighting each diagnostic ester carbonyl (C-9) correlation with its glucosyl-1 or -2-proton, delineating the two independent sinapoyl units attached to the 1- (S1) vs 2-position (S2) of glucose. C, the corresponding structure of 1,2-disinapoylglucose, compound 2.
**Figure 5.** Analysis of sinapate esters in mature leaves.

Extracts were prepared from leaves of 5-week-old wild-type, *sng1-1, sng1-5,* and *sng1-6* plants and were analyzed by HPLC using UV detection at 335 nm. SG, sinapoylglucose, SM, sinapoylmalate, 1,2-DSG, 1,2-disinapoylglucose. 1,2-disinapoylglucose accumulates in the leaves of only the *sng1-1* and *sng1-5* mutants.

**Figure 6.** HPLC analysis of soluble secondary metabolites accumulated in the leaves of wild type, *brt1-1, sng1-1* and the *brt1-1/sng1-1* double mutant.

Leaves were extracted with 50% MeOH and analyzed by HPLC. The elution of UV-absorbing compounds was monitored at 322 nm. SG, sinapoylglucose; SM, sinapoylmalate; 1,2-DSG, 1,2-disinapoylglucose. Col, wild-type Columbia. 1,2-disinapoylglucose accumulation is absolutely dependent upon the activity of the BRT1 glycosyltransferase.

**Figure 7.** Analysis of anthocyanins in 11-day-old, sucrose-stressed seedlings.

Acidic methanol extracts were prepared from 11-day-old Columbia wild-type, *sng1-1, sng1-5,* and *sng1-6* seedlings grown for six days in liquid MS media containing 1% sucrose, and for five additional days in MS media containing 10%...
sucrose. Extracts were analyzed by HPLC using detection at 525 nm. Peaks labeled A9 and A11 indicate the sinapoylated anthocyanins, 3-O-[6-O-p-coumaroyl-2-O-(2-O-sinapoyl-β-D-xylopyranosyl)-β-D-glucopyranosyl]-5-O-(6-O-malonyl-β-D-glucopyranosyl)cyanidin and 3-O-[6-O-(4-O-β-D-glucopyranosyl-p-coumaroyl)-2-O-(2-O-sinapoyl-β-D-xylopyranosyl)-β-D-glucopyranosyl]-5-O-(6-O-malonyl-β-D-glucopyranosyl)cyanidin, respectively. A5 and A8 indicate non-sinapoylated anthocyanins 3-O-(6-O-p-coumaroyl-2-O-β-D-xylopyranosyl-β-D-glucopyranosyl)-5-O-(6-O-malonyl-β-D-glucopyranosyl)cyanidin and 3-O-[6-O-(4-O-β-D-glucopyranosyl-p-coumaroyl)-2-O-β-D-xylopyranosyl-β-D-glucopyranosyl]-5-O-(6-O-malonyl-β-D-glucopyranosyl)cyanidin, respectively (Tohge et al. 2005).

**Figure 8.** Analysis of the sng1-5/SNG1 transformant.

Acidic methanol extracts were prepared from 11-day-old, sucrose-stressed Columbia wild-type and sng1-5 transformed with the At2g22990/SMT genomic construct (pCC398) and analyzed by HPLC using detection at 335 nm and 525 nm. Both wild-type and sng1-5/pCC398 seedlings accumulate sinapoylmalate (SM) and non-sinapoylated anthocyanins (A5, A8; 3-O-(6-O-p-coumaroyl-2-O-β-D-xylopyranosyl-β-D-glucopyranosyl)-5-O-(6-O-malonyl-β-D-glucopyranosyl)cyanidin and 3-O-[6-O-(4-O-β-D-glucopyranosyl-p-coumaroyl)-2-O-β-D-xylopyranosyl-β-D-glucopyranosyl]-5-O-(6-O-malonyl-β-D-glucopyranosyl)cyanidin, respectively), but only the wild type accumulates sinapoylated anthocyanins (A9 and A11; 3-O-[6-O-p-coumaroyl-2-O-(2-O-
sinapoyl-β-D-xylopyranosyl)-β-D-glucopyranosyl]-5-O-(6-O-malonyl-β-D-glucopyranosyl)cyanidin and 3-O-[6-O-(4-O-β-D-glucopyranosyl-p-coumaroyl)-2-O-(2-O-sinapoyl-β-D-xylopyranosyl)-β-D-glucopyranosyl]-5-O-(6-O-malonyl-β-D-glucopyranosyl)cyanidin, respectively).

**Figure 9.** Possible metabolic routes of anthocyanin acylation in Arabidopsis.

Anthocyanins **A1-A11** have been identified in wild-type Arabidopsis and the *pap1-D* mutant (Bloor and Abrams, 2002; Tohge et al. 2005). Arrows indicate potential acyltransferase reactions leading to the addition of malonyl (M), *p*-coumaroyl (*p*-C) and sinapoyl (S) groups, and glycosylation of the *p*-coumaroyl group (Glc). **A1**, 3-O-(2-0-β-D-xylopyranosyl-β-D-glucopyranosyl)-5-O-β-D-glucopyranosyl-cyanidin; **A2**, 3-O-(2-O-β-D-xylopyranosyl-β-D-glucopyranosyl)-5-O-(6-O-malonyl-β-D-glucopyranosyl)cyanidin; **A3**, 3-O-(6-O-p-coumaroyl-2-O-β-D-xylopyranosyl-β-D-glucopyranosyl)-5-O-β-D-glucopyranosyl-cyanidin; **A4**, 3-O-[2-O-(2-O-sinapoyl-β-D-xylopyranosyl)-β-D-glucopyranosyl]-5-O-β-D-glucopyranosyl-cyanidin; **A5**, 3-O-(6-O-p-coumaroyl-2-O-β-D-xylopyranosyl-β-D-glucopyranosyl)-5-O-(6-O-malonyl-β-D-glucopyranosyl)cyanidin; **A6**, 3-O-[6-O-(4-O-β-D-glucopyranosyl-p-coumaroyl)-2-O-β-D-xylopyranosyl-β-D-glucopyranosyl]-5-O-β-D-glucopyranosyl-cyanidin; **A7**, 3-O-[6-O-p-coumaroyl-2-O-(2-O-sinapoyl-β-D-xylopyranosyl)-β-D-glucopyranosyl]-5-O-β-D-glucopyranosyl-cyanidin; **A8**, 3-O-[6-O-(4-O-β-D-glucopyranosyl-p-coumaroyl)-2-O-β-D-xylopyranosyl-β-D-glucopyranosyl]-5-O-(6-O-malonyl-β-D-glucopyranosyl)cyanidin; **A9**, 3-O-[6-O-p-
coumaroyl-2-O-(2-O-sinapoyl-β-D-xylopyranosyl)-β-D-glucopyranosyl]-5-O-(6-O-malonyl-β-D-glucopyranosyl)cyanidin; A10, 3-O-[6-O-(4-O-β-D-glucopyranosyl-p-coumaroyl)-2-O-(2-O-sinapoyl-β-D-xylopyranosyl)-β-D-glucopyranosyl]-5-O-β-D-glucopyranosylcyalinidin; A11, 3-O-[6-O-(4-O-β-D-glucopyranosyl-p-coumaroyl)-2-O-(2-O-sinapoyl-β-D-xylopyranosyl)-β-D-glucopyranosyl]-5-O-(6-O-malonyl-β-D-glucopyranosyl)cyanidin. 3-O-[2-O-(2-O-Sinapoyl-β-D-xylopyranosyl)-β-D-glucopyranosyl]-5-O-(6-O-malonyl-β-D-glucopyranosyl)cyanidin is designated as X since it has not been identified in Arabidopsis to date and, as a result, the lines that lead to and from X are not shown with arrowheads since it is not clear that it is a genuine biosynthetic intermediate in anthocyanin biosynthesis.
cyanidin 3-O-[6-O-(4-O-β-D-glucopyranosyl-p-coumaroyl)-2-O-(2-O-
sinapoyl-β-D-xylopyranosyl)-β-D-glucopyranosyl]-5-O-(6-O-
malonoyl-β-D-glucopyranoside)
At2g22970
At2g22980
At2g22990
At2g23000
At2g23010
At2g23020
At2g23030
At2g23040
At2g23050
At2g23060
At2g23070

sng1-1
SALK_133207

sng1-5

sng1-6
Absorbance at 335nm (mAU)

- **wild type**
- **sng1-1**
- **sng1-5**
- **sng1-6**

Time (min): 0, 5, 10, 15, 20, 25, 30

Fraser et al. Figure 5
Fraser et al. Figure 6

Absorbance at 322 nm

Retention time (min)

Col

brt1-1

sng1-1

1,2-DSG

brt1-1/sng1-1

Retention time (min)
Fraser et al. Figure 7