PERGAMON

Phytochemistry 59 (2002) 343–346

PHYTOCHEMISTRY

www.elsevier.com/locate/phytochem

The structure of the major anthocyanin in *Arabidopsis thaliana*

Stephen J. Bloor[a,*], Sharon Abrahams[b]

[a]*Industrial Research Ltd., PO Box 31-310, Lower Hutt, New Zealand*
[b]*CSIRO Plant Industry, PO Box 1600, Canberra ACT, 2601, Australia*

Received 6 August 2001; received in revised form 6 November 2001

Abstract

The major anthocyanin in the leaves and stems of *Arabidopsis thaliana* has been isolated and shown to be cyanidin 3-*O*-[2-*O*(2-*O*-(sinapoyl)-β-D-xylopyranosyl)-6-*O*-(4-*O*-(β-D-glucopyranosyl)-*p*-coumaroyl-β-D-glucopyranoside] 5-*O*-[6-*O*-(malonyl) β-D-glucopyranoside]. This anthocyanin is a glucosylated version of one of the anthocyanins found in the flowers of the closely related *Matthiola incana*. © 2002 Elsevier Science Ltd. All rights reserved.

Keywords: Arabidopsis thaliana; Brassicaceae; Anthocyanin; Acylated cyanidin 3-sambubioside-5-glucoside; Malonic acid; Sinapic acid; *p*-Coumaric acid

1. Introduction

The metabolic pathway in plants leading to the flavonoids is one of the most intensively studied secondary metabolite systems. *Arabidopsis thaliana* L. (thale cress, Brassicaceae) is a highly favoured model for the study of this pathway owing to the ready availability of a wide range of mutants with altered flavonoid biosynthesis (the transparent testa (*tt*) mutants) (Kornneef, 1990; Graham, 1998). Despite the widespread use of *Arabidopsis* for this purpose it is only recently that the precise structures of some of the flavonoid products themselves have been described. Veit and Pauli (1999) determined the structures of the major flavonols in *Arabidopsis* leaf tissue. Our interest lies in the other major end-products of the pathway in *Arabidopsis*, the proanthocyanidins and the anthocyanins. In this communication we describe the major anthocyanin present in the leaves and stems of mature wild type plants.

2. Results and discussion

Plants grown in growth chambers produce variable amounts of anthocyanins in their leaves and stems dependent upon the light level available. The major anthocyanin **(1)** was purified from an acidic aqueous leaf extract by RP CC and Sephadex LH20 CC. A second anthocyanin **(2)** was similarly purified from a larger scale acidic MeOH extract of the stems. Compound **2** was later shown to be a methylated derivative of the native anthocyanin **(1)** produced in the extraction process. As **2** was purified in greater quantity most of the structural work was performed on this compound.

The ^{1}H NMR spectrum of **2** showed signals for a cyanidin aglycone, four β-sugars, a sinapic acid and a *p*-coumaric acid group. Additionally, the ^{13}C NMR spectrum showed the typical signals of a malonic acid group as well as two methoxyl resonances. In combination with the ESMS, which gave a M^+ at 1342 amu, these data indicated **2** to be made up of a cyanidin, 3 hexose sugars, 1 pentose sugar, the aforementioned sinapic, *p*-coumaric and malonic acids as well as the extra methyl group.

Prior to undertaking more detailed spectral studies of **1** and **2**, a review of data reported for anthocyanins from other Brassica species suggested the *Arabidopsis* anthocyanin might be related to the *Matthiola incana* anthocyanins (Saito et al., 1995). Indeed, HPLC comparison of an extract of red/violet *Matthiola* petals tissue and **1** showed it and the major *Matthiola* anthocyanin had similar on-line UV-vis spectra but that their retention times were considerably different. More importantly, both anthocyanins yielded the same cyanidin glycoside after alkaline hydrolysis. Thus, the

---

[*] Corresponding author. Tel.: +64-4-569-0576; fax: +64-4-569-0055.

E-mail addresses: s.bloor@irl.cri.nz (S.J. Bloor), s.abrahams@pi.csiro.au (S. Abrahams).

PII: S0031-9422(01)00460-5

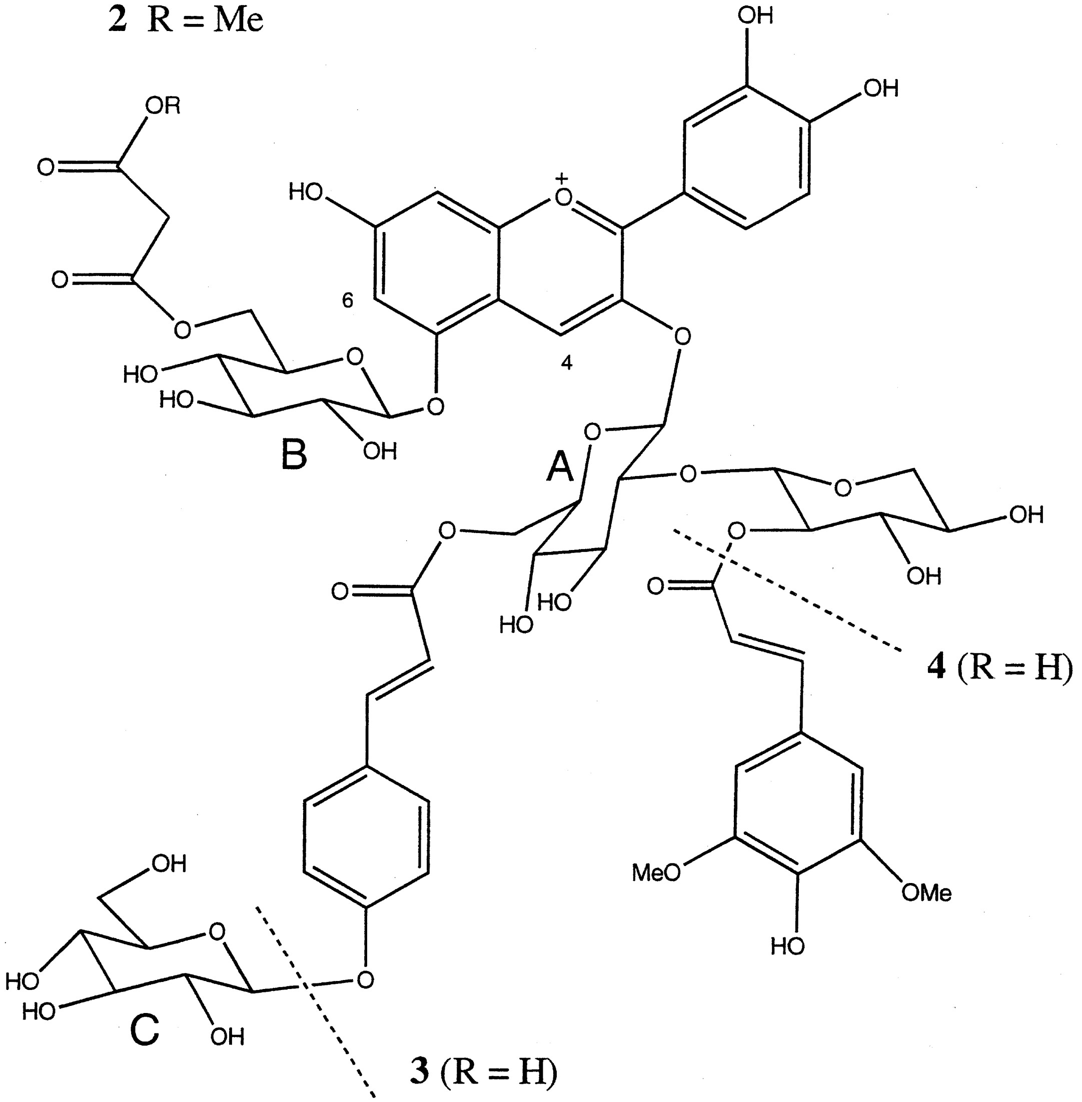
basic unacylated cyanidin glycoside in **1** and **2** is the same as that in the red/violet *Matthiola* flowers, i.e. cyanidin 3-(xylosyl(1→2)glucoside)-5-glucoside.

Further evidence to complete the structure was gained from a series of NMR experiments on **2**. Firstly, the anomeric proton signal at 5.6 ppm (GlcA) could be linked to the cyanidin C-3 (HMBC) and cyanidin H-4 (ROESY). The relative downfield shift of the ^{13}C and ^{1}H NMR signals of the H-2 of GlcA and a correlation of the signals with the xylose H-1 (HMBC) supported the conclusion from above that the disaccharide was a sambubiose. The two aromatic acyl groups could be placed on the C-6 of GlcA (*p*-coumaric acid) and the C-2 of the xylose (sinapic acid) due to the appearance of long range C-H coupling between the respective sugar protons and the acyl carbonyl carbons. Other connections, made using similar arguments, defined one of the overlapping anomeric proton signals at 5.2 ppm to be that of the β-D-glucose substituent (GlcB) at C-5 of the cyanidin. The H-6 protons of GlcB showed long range correlation (HMBC) to one of the malonic acid carbonyl signals. The other malonic acid carbonyl shows similar correlation to a MeO signal showing **2** is a methyl ester. Finally the remaining glucose (GlcC) could be attached at the C-4 position of the *p*-coumaroyl group (the GlcA anomeric proton demonstrated a long range correlation with the characteristic ^{13}C NMR signal of the *p*-coumaroyl C-4 at 161 ppm).

Thus **2** is cyanidin 3-*O*-[2-*O*-(2-*O*-(sinapoyl)-β-D-xylopyranosyl)-6-*O*-(4-*O*-(β-D-glucopyranosyl)-*p*-coumaroyl)-β-D-glucopyranoside] 5-*O*-[6-*O*-(malonyl) β-D-glucopyranoside] methyl ester. Hence, the native anthocyanin **(1)** is clearly the free malonic acid. The ESMS of **1** has the expected M^+ at 1343 amu. The ^{1}H NMR spectrum of **1** is virtually identical to that of **2**, apart from the absence of the MeO signal at 3.64 ppm.

1 R = H

2 R = Me



4 (R = H)

3 (R = H)

Methylation and subsequent loss of malonic acid groups in anthocyanins acylated with this acid is well documented. When extracts of the stem were left in acidic MeOH at RT a slow methylation took place, followed by loss of the malonic acid group altogether.

Compound **1** is the dominant anthocyanin in both the leaves and stems of mature wild type (WT) plants. Among the minor anthocyanins seen in the HPLC chromatogram of the extracts (see Fig. 1), two other anthocyanins are of interest. In the leaf extract, a significant peak at R_t 14 mins, **4**, was assigned to be a desinapoylated version of **1**. This anthocyanin has a much lower absorption in the 300–320 nm region of the hplc on-line UV-vis spectrum and the ESMS of a partially purified fraction shows a M^+ at 1137 amu consistent with this. The extract of the stems and siliques had very little of anthocyanin **4** present but the biggest minor peak was a later eluting peak, **3** (see Fig. 1). This anthocyanin had an identical on-line UV-vis spectrum to that of **1** but the ESMS of a partially purified fraction indicated a molecular weight of 1181 amu, and so could be assigned to be the structure shown for **3**, i.e. no terminal glucose. This latter structure has been previously described from *Matthiola incana* flowers (Saito et al., 1995). During our investigations we have also examined the anthocyanins from two of the *Arabidopsis tt* mutants; the *tt*10 mutant is similar to WT, i.e. **1** is the major leaf anthocyanin (from HPLC comparison and ESMS of purified anthocyanin) whereas the leaves of the *tt*7 mutant have the pelargonidin equivalent of **1** (HPLC on-line spectrum has vis. max at 518 nm, ESMS 1327 amu).

Analysis of the three major acyanic flavonoids present confirms the earlier reports of kaempferol 7-*O*-rhamnosides with a 3-*O*-rhamnose,-glucose or-disaccharide combination (Veit and Pauli, 1999). In our extracts the disaccharide moiety was a neohesperidoside (rhamnosyl(1→2)glucoside) rather than the gentiobioside (glucosyl(1→6)glucoside) reported by Veit and Pauli (1999).

3. Experimental

3.1. General procedures

NMR experiments were run at 500 MHz or at 300 MHz (75 MHz for ^{13}C). Anthocyanin samples were dissolved in 2% CF_3COOD in CD_3OD. Flavonols were run in DMSO-d_6. MS were obtained using a VG Quattro (Electrospray). RP HPLC analyses were performed using a Waters 2690 solvent delivery system coupled to a Waters 996 PDA detector. The column (Merck Purosphere STAR (55×4 mm, 3 μm) was eluted with a solvent system comprising 95:5 H_2O:HCOOH (solvent A) and MeOH mixed using a linear gradient starting with 90% A, decreasing to 78% A after 15 min, 50% A after 35 min, 30% A after 38 min and 5% A after 40 min.

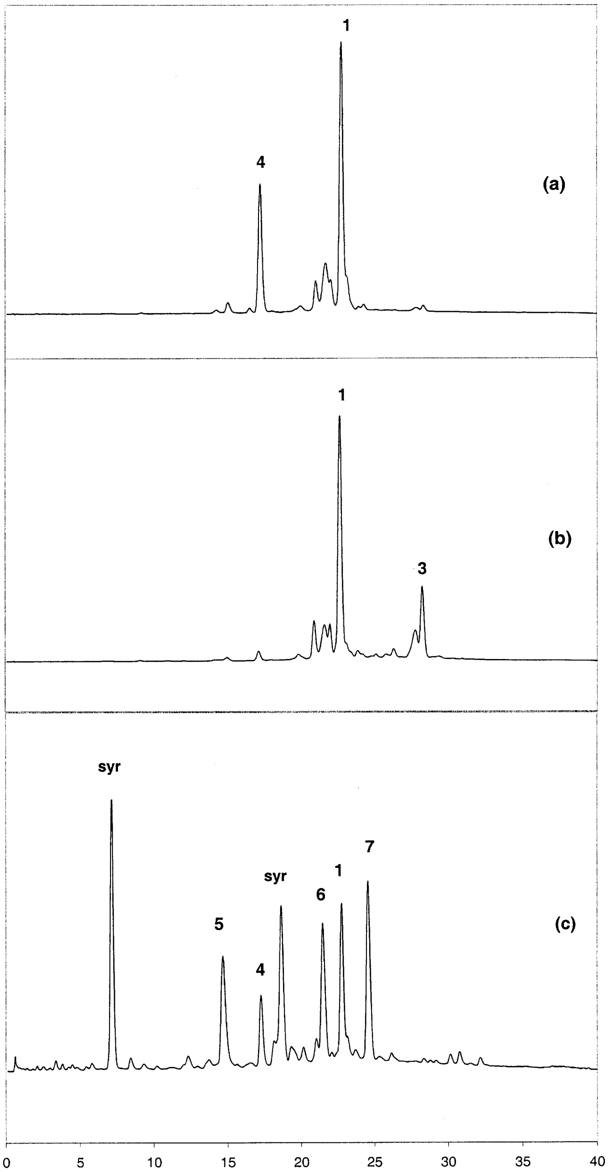


Fig. 1. HPLC chromatograms (abs. vs time(mins)) of (a) leaf extract at 520 nm, (b) stem/silique extract at 520 nm and (c) leaf extract at 340 nm. Numbers correspond to compounds described in the text. **1**. Cyanidin 3-*O*-[2-*O*-(2-*O*-(sinapoyl)-xylosyl)-6-*O*-(4-*O*-(glucosyl)-*p*-coumaroyl)) glucoside] 5-*O*-[6-*O*-(malonyl) glucoside], **3**. Cyanidin 3-*O*-[2-*O*-(2-*O*-(sinapoyl) xylosyl)-6-*O*-(*p*-coumaroyl) glucoside] 5-*O*-[6-*O*-(malonyl) glucoside], **4**. Cyanidin 3-*O*-[2-*O*-(xylosyl)-6-*O*-(4-*O*-(glucosyl)-*p*-coumaroyl)) glucoside] 5-*O*-[6-*O*-(malonyl) glucoside], **5**. Kaempferol 3-*O*-[rhamnosyl(1→2) glucoside]-7-*O*-rhamnoside, **6**. Kaempferol 3-*O*-glucoside-7-*O*-rhamnoside, **7**. Kaempferol 3-O-rhamnoside-7-O-rhamnoside. Syringic acid esters are denoted by syr (assigned from on-line spectra).

3.2. Isolation of anthocyanins and flavonols

Plants were grown in growth cabinets under 16 h day at 24 °C and 8 h night at 18 °C. *Arabidopsis thaliana* L.

seeds [WT ecotype Columbia 7 (cs 3731)] were obtained from the Arabidopsis Biological Resource Centre (Ohio State University, USA). Fresh or frozen plant tissue from mature plants was extracted by pulverizing in 0.2% aq. TFA or HCOOH:MeOH (3:97). The volume of the filtered extract was reduced to a minimum volume of water and the solution applied directly to a RP column. The column was eluted with an acidic-water: methanol gradient. The fractions containing the major pigment and flavonols were further purified by Sephadex LH-20 CC (MeOH:TFA:H_2O, 500:2:498) to yield **1** from the aq. extract or **2** from the acidic MeOH extract and the three flavonol glycosides.

3.3. Cyanidin 3-O-[2-O-(2-O-(sinapoyl)-β-D-xylopyranosyl)-6-O-(4-O-(β-D-glucopyranosyl)-p-coumaroyl)- β-D-glucopyranoside] 5-O-[6-O-(malonyl) β-D-glucopyranoside], ***1***

UV-vis (HPLC on-line) 284, 296, 320 (*sh*), 539 nm. LR ESMS 1343 ($C_{61}H_{67}O_{34}$ = 1343.35). ^{1}H NMR data: as described for **1** in Table 1, but without methyl signal at 3.64 ppm.

Table 1
^{13}C and ^{1}H NMR data for **2**, in CD_3OD/CF_3COOD (98:2), δ in ppm

		^{13}C	^{1}H
Cyanidin	2	164.9	
	3	146.3	
	4	133.8	8.71 *s*
	5	156.9	
	6	106.4	6.91 *d* (1.5)
	7	169.5	
	8	97.8	6.79 *d* (1.5)
	9	156.7	
	10	113.3	
	1	121.6	
	2	117.0	7.93 *d* (2.5)
	3	148.2	
	4	157.4	
	5	117.9	7.01 *d* (8.5)
	6	129.8	8.35 *dd* (8.5, 2.5)
3-glc	1	100.2	5.61 *d* (7.5)
	2	79.3	4.07 *t* (7.5)
	6	65.0	4.39, 4.47
5-glc	1	102.1	5.18 *d* (7.5)
	6	65.5	4.26, 4.49
xyl	1	102.5	5.18 *d* (7.5)
	2	75.7	4.82 *t* (8.5)
	5	67.7	3.30, 4.02
Term-glc	1	102.1	4.98 obsc.
	6	63.2	3.7–3.9

^{13}C signals for acyl groups in **2:** *p*-coumaric acid group; 169.1 (C = O), 117.0 (C-α), 146.6 (C-β), 129.8 (C-1), 131.5 (C-2,6), 118.2 (C-3,5) and 161.3 (C-4); .sinapic acid 168.9 (C = O), 116.4 (C-α), 147.8 (C-β), 126.9 (C-1), 107.3 (C-2,6), 149.9 (C-3,5),140.2 (C-4), 57.4 (OMe); methyl malonic acid 168.7 (inner C = O), 169.3 (outer C = O), 42.1 (CH_2), 53.5 (OMe).

3.4. Cyanidin 3-O-[2-O-(2-O-(sinapoyl)-β-D-xylopyranosyl)-6-O-(4-O-(β-D-glucopyranosyl)-p-coumaroyl)-β-D-glucopyranoside] 5-O-[6-O-(malonyl) β-D-glucopyranoside] methyl ester, ***2***

UV-vis (HPLC on-line) 284, 296, 320 (sh), 539 nm. LR ESMS 1358 ($C_{62}H_{69}O_{34}$ = 1357.37). ^{1}H and ^{13}C NMR data: see Table 1.

3.5. Kaempferol 3-O-β-[β-D-rhamnopyranosyl(1→2)D-glucopyranoside]-7-O-α-L-rhamnopyranoside, ***5***

UV, MS, ^{13}C and ^{1}H NMR data were in agreement with literature values (Mulinacci et al., 1995).

3.6. Kaempferol 3-O-β-D-glucopyranoside-7-O-α-L-rhamnopyranoside (6), and kaempferol 3-O-α-L-rhamnopyranoside-7-O-α-L-rhamnopyranoside (7).

UV, MS and ^{1}H NMR data were in agreement with literature values (Pauli, 2000).

Acknowledgements

The authors thank the Department of Chemistry, Australian National University, ACT, Australia for a visiting fellowship (S.J.B.). This work was funded by the New Zealand Foundation for Research Science and Technology (S.J.B., Contract CO-8X0004) and Pioneer Hybrid (SLA).

References

Graham, T.L., 1998. Flavonoid and flavonol glycoside metabolism in *Arabidopsis*. Plant Physiol. Biochem. 36, 135–144.

Kornneef, M., 1990. Mutations affecting the testa colour in *Arabidopsis*. Arabidopsis Inf. Ser. 27, 1–4.

Mulinacci, N., Vincieri, F.F., Baldi, A., Bambagiotti-Alberti, M., Sendl, A., Wagner, H., 1995. Flavonol glycosides from *Sedum telephium* subspecies *maximum* leaves. Phytochemistry 38, 531–533.

Pauli, G.F., 2000. Higher order and substituent chemical shift effects in the proton NMR of glycosides. J. Nat. Prod. 63, 834–838.

Saito, N., Tatsuzawa, F., Nishiyama, A., Yokoi, M., Shigihara, A., Honda, T., 1995. Acylated cyanidin 3-sambubioside-5-glucosides in *Matthiola incana*. Phytochemistry 38, 1027–1032.

Veit, M., Pauli, G.F., 1999. Major flavonoids from *Arabidopsis thaliana* leaves. J. Nat. Prod. 62, 1301–1303.