Structural Characterization of the Major Flavonoid Glycosides from Arabidopsis thaliana Seeds

ARTICLE *in* JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY · OCTOBER 2006

Impact Factor: 2.91 · DOI: 10.1021/jf061043n · Source: PubMed

7 AUTHORS, INCLUDING:

Jean-Marc Routaboul
French National Institute for Agricultural Res…
27 PUBLICATIONS **2,038** CITATIONS
SEE PROFILE

Loic Lepiniec
French National Institute for Agricultural Res…
101 PUBLICATIONS **7,307** CITATIONS
SEE PROFILE

Nicolas Birlirakis
Ecole Normale Supérieure de Paris
64 PUBLICATIONS **832** CITATIONS
SEE PROFILE

JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

J. Agric. Food Chem. **2006,** *54,* 6603−6612 **6603**

Structural Characterization of the Major Flavonoid Glycosides from *Arabidopsis thaliana* Seeds

LUCIEN KERHOAS,[†] DENYA AOUAK,[†] ANNABELLE CINGÖZ,[†] JEAN-MARC ROUTABOUL,[‡] LOÏC LEPINIEC,[‡] JACQUES EINHORN,*[,†] AND NICOLAS BIRLIRAKIS[§]

Unité de Phytopharmacie et Médiateurs Chimiques and Laboratoire de Biologie des Semences, INRA, route de St Cyr, 78026 Versailles cedex, France; and Laboratoire de RMN à Haut Champ, ICSN-CNRS, 91198 Gif-sur-Yvette, France

Information gains from the seed of the model plant *Arabidopsis thaliana* (Brassicaceae) have greatly contributed to a better understanding of flavonoid synthesis and may be used for crop improvement. However, exhaustive identification of the flavonoid accumulated in *Arabidopsis* seed was still lacking. Complementary investigations of seed flavonoids by LC-ESI-MS, LC-ESI-MS-MS, and NMR spectroscopy in *Arabidopsis* led to full characterization of monoglycosides, namely, quercetin 3-*O*-α-rhamnopyranoside and kaempferol 3-*O*-α-rhamnopyranoside, and diglycosides, namely, quercetin and kaempferol 3-*O*-β-glucopyranoside-7-*O*-α-rhamnopyranoside and quercetin and kaempferol 3,7-di-*O*-α-rhamnopyranoside. Interestingly, the *tt7* mutant that lacks flavonoid-3′-hydroxylase and consequently accumulates only kaempferol derivatives was shown to contain three additional products, kaempferol 3-*O*-β-glucopyranoside, kaempferol 3-*O*-α-rhamnopyranoside-7-*O*-β-glucopyranoside, and the triglycoside kaempferol 3-*O*-β-[α-rhamnopyranosyl(1→2)-glucopyranoside]-7-*O*-α-rhamnopyranoside. Taken together these results allow a scheme for flavonol glycosylation to be proposed.

KEYWORDS: ***Arabidopsis thaliana*** **seeds; LC-ESI-MS-MS; NMR; quercetin glycosides; kaempferol glycosides; flavonoids; glycosylation process**

INTRODUCTION

Understanding basic processes in *Arabidopsis thaliana* as a model plant will facilitate the genetic engineering of crops to introduce value-added nutraceuticals or improve forage quality traits. Indeed, because higher plant genomes are closely related, the knowledge acquired on *Arabidopsis* may be transferred to agronomic plants (*1*−*3*). *A. thaliana* L. (thale cress; Brassicaceae) has become a model in the study of the biosynthesis and regulation of the flavonoid pathway due to the availability of a wide range of mutants affected in this pathway (*4*). They have contributed to a careful dissection of the roles, biochemistry, and molecular biology of flavonoids (*5*, *6*). For instance, the *tt7* mutant does not contain quercetin derivatives because the gene coding for the flavonoid-3′-hydroxylase enzyme that synthesizes dihydroquercetin from dihydrokaempferol is deficient (*7*). This mutant shows reduced auxin transport (*8*) and lower resistance to UV-B radiation (*9*).

Flavonoids are an integral part of human and animal diets. For example, apples, beans, berries, cereals, citrus, grape, onions, and tea constitute rich sources of flavonoids. These compounds are responsible for major organoleptic characteristics and have impacts on many agronomical traits (*5*, *6*). High concentrations of astringent tannins can have a negative impact on the nutritive value and palatability of forage. On the other hand, the presence of tannins prevents pasture bloat, which can be lethal for the animal (*10*, *11*). Laboratory and epidemiological studies suggest a beneficial effect of flavonoids, preventing the occurrence of chronic age-related diseases such as cardiovascular functions (*12*, *13*) or some cancers (*14*). These plant quality traits explain why there is a growing interest in a better understanding of flavonoid accumulation and biosynthesis regulation.

Flavonoids can be classified into several classes depending on the degree of unsaturation and oxidation of the basic 15-carbon skeleton, but flavonols represent the most important group (*15*, *16*). For instance, more than 350 derivatives evolving from a single flavonol aglycone, quercetin, have been described in plants (*15*). Most of this variability comes from glycosylation by one or several sugar moieties. Mass spectrometric analysis techniques coupled to liquid chromatography (LC-ESI-MS) are well suited for flavonoid analysis, providing the molecular weight, the general structure, and relative concentrations of the various constituents with high sensitivity and specificity (*17*). In-depth structural information on several features can be obtained subsequently from MS-MS experiments performed in

* Author to whom correspondence should be addressed [telephone + 33 (0)1 30 83 31 20; fax + 33 (0)1 30 83 31 19; e-mail einhorn@versailles.inra.fr].

[†] Unité de Phytopharmacie et Médiateurs Chimiques, Institut National de la Recherche Agronomique (INRA).

[‡] Laboratoire de Biologie des Semences, INRA.

[§] Centre National de la Recherche Scientifique (CNRS).

10.1021/jf061043n CCC: $33.50
Published on Web 08/10/2006

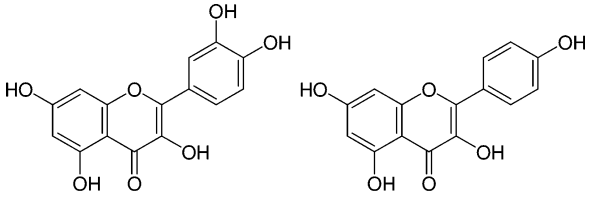


Figure 1. Structures of quercetin and kaempferol aglycones.

the positive or negative mode (*18*, *19*). Moreover, high-resolution NMR spectroscopy permits validation of hypotheses made and provides further insight into fine structural aspects inaccessible by MS.

Surprisingly, the structures of some of the flavonols that accumulate in *Arabidopsis* leaves and seed were only revealed recently (*20*−*24*). In contrast to *Arabidopsis* leaves, which essentially synthesize kaempferol derivatives (**Figure 1**), seeds were found to mainly accumulate quercetin glycosides and procyanidins (*24*). In fact, limited information was provided on the structures of the flavonoid glycosides. The aim of this study was thus to achieve the isolation and complete identification of the main flavonols found in seeds of *A. thaliana* L. (ecotype Ws-2) by combining LC-ESI-MS-MS and NMR investigations. The flavonoid composition of the flavonoid-3′-hydroxylase deficient *tt-7* mutant seed has also been investigated for comparison and to characterize the less abundant kaempferol glycosides present in Ws-2. Differences regarding the glycosylation scheme between *Arabidopsis* seed and leaves are discussed. Our results show that the flavonol composition of *Arabidopsis* seed is representative of other seeds and grains.

MATERIALS AND METHODS

Plant Material. *A. thaliana* (L.) Heynh, Wassilewskija-2 (Ws-2) and *tt7-4* (DJI11), from the INRA-Versailles collection have been described previously (*20*). Plants were grown in a controlled growth chamber with a 16 h photoperiod (20 °C/15 °C, light/dark, 60% humidity under 100 μEinstein/m^2/s white light).

Extraction and Isolation of Flavonoids. *Extraction Procedure for Profiling and LC-MS Studies.* Fifteen milligrams of seeds was thoroughly ground in a 10 mL Potter-Elvehjem grinder (VWR, Fontenay-sous-Bois, France) with 4 mL of acetonitrile/water (75:25, v/v) for 5 min at 4 °C. A 6 μg sample of apigenin was introduced at the beginning of the extraction as internal standard. Two milliliters of ultrapure (UP) water was added, and extraction was completed by sonication for 30 min. After filtration through a 0.45 μm PVDF membrane filter (Interchim), the crude extract was directly subjected to LC-MS analysis.

Extraction and Isolation Procedures of the Major Seed Flavonoids. A scaled up protocol was applied for flavonoid isolation using a total of 10 g of seeds of either Ws-2 or *tt7-4* mutant material: 2 g subsamples were ground with 10 mL of acetonitrile/water (75:25, v/v) in a glass Duall grinder and sonicated for 20 min at 4 °C. After centrifugation, the pellets were re-extracted twice under 20 min of sonication, and the combined extracts were evaporated under reduced pressure to eliminate acetonitrile. The total extract (10 g seeds) was diluted with UP water to ≈50 mL and delipidated three times with 5 mL of *n*-hexane washes. Purification using cation and anion exchange cartridges successively was done to eliminate coextracted non-flavonoid compounds retained specifically on either support. The pH of the aqueous extract was adjusted to 2.5 (0.1 N HCl), and fractions of 5 mL were purified on 400 mg Chromafix PS-H cartridges (Macherey-Nagel, Hoerdt, France). The resulting acidic extract was neutralized (0.1 N NaOH) and then diluted to 500 mL. Aliquots of 50 mL were applied for further purification to 1 g SAX Bond Elut cartridges (Varian, Les Ulis, France). After lyophilization, the residue was dissolved in methanol/water and fractionated by semipreparative HPLC using a 250 × 10.5 mm i.d., 5 μm, Kromasil C18 column (Merck-Clévenot, Nogent-sur-Marne, France). A gradient of acetonitrile (solvent A) and UP water (solvent B) was applied at a 3.5 mL/min flow rate: 0−5 min, 10% A; 5−30 min, 30% A (linear gradient); 30−31 min, 100% A. The column was washed with 100% A for 15 min and equilibrated for 15 min at 10% A between runs. Elution was monitored by UV at 280 nm and MS (0.1 mL/min split), and the flavonoids were manually collected after the UV detector. All fractions were concentrated under reduced pressure prior lyophilization.

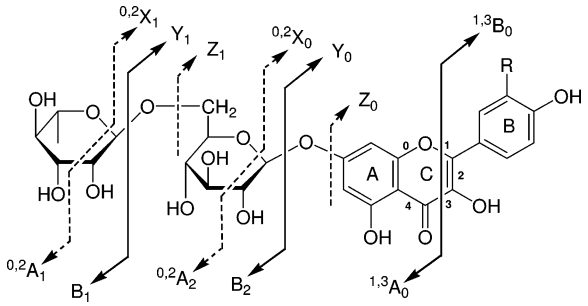


Figure 2. Ion nomenclature for MS-MS spectra of flavonoid glycosides.

Mass Spectrometry. *Instrumental Conditions.* Mass analysis was conducted on aliquots of the final extracts or purified fractions using a Quattro LC instrument with an ESI "Z-spray" interface (Micromass, Manchester, U.K.), MassLynx software, an Alliance 2695 separation module (Waters, Milford, MA), and a Waters 2487 dual UV detector. Separation was achieved on a 150 × 2 mm i.d., 5 μm, reverse-phase Uptisphere C_{18} column (Interchrom, Montluçon, France) using a mix comprising solvent A, acetonitrile/water (95:5, v/v, 0.5% acetic acid), and solvent B, acetonitrile/water (5:95, v/v, 0.5% acetic acid), with a gradient profile (starting with 10% A for 5 min; linear gradient to 70% A over 35 min; a washing step 100% A for 15 min, and final equilibration at 10% A for 20 min) and 0.2 mL/min flow rate. Column temperature was maintained at 24 °C, and UV detection was set at 280 nm. The ESI source potentials were as follows: capillary, 3.25 kV (positive mode) or 3.0 kV (negative mode); extractor, 2 V. The source block and desolvation gas were heated at 120 and 400 °C, respectively. Nitrogen was used as nebulization and desolvation gas (100 and 500 L/h, respectively). Argon as collision gas was set at 3.5 × 10^{-3} mbar for the collision-induced dissociation (CID) MS-MS experiments (10−50 eV collision energy).

Ion Nomenclature. The product ions issuing from the MS-MS experiments are named according to the nomenclature proposed by Domon and Costello (*25*) (**Figure 2**). Ions containing the aglycone are labeled $^{k,l}X_j$, Y_j, and Z_j, where j is the number of the interglycosidic bond broken, counting from the aglycone, and the superscripts k and l indicate the cleavages within the carbohydrate rings. If the charge is retained on the carbohydrate residue, product ions are named A_i, B_i, or C_i, where i represents the number of the glycosidic bond cleaved from the nonreducing terminus. For a free aglycone, the ^{i,j}A and ^{i,j}B labels refer, as proposed by Ma et al. (*26*), to the product ions containing intact A- and B-rings, respectively, in which the superscripts i and j indicate the CC or CO-ring bonds that have been broken.

NMR Spectroscopy. NMR experiments were carried out at 298 K on a Bruker Avance 800 spectrometer using an TXI cryoprobe, equipped with z gradients. Spectra were processed with TOPSPIN software. Perdeuterated methanol was used as solvent and was purchased from Eurisotop (Saint-Aubin, France). Homonuclear and heteronuclear experiments were acquired with standard pulse sequences; a 70 ms DIPSI2 spin−lock pulse was used for TOCSY mixing, and broadband adiabatic decoupling was applied during the acquisition time of the ^{1}H−^{13}C HSQC experiment. The low-pass filter (70 ms) version of the HMBC sequence was used to attenuate one-bond correlations.

RESULTS AND DISCUSSION

Characterization of the Main Flavonoids Using LC-MS-MS. The ESI-LC-MS analyses allowed the first structure hypotheses to be established. The use of the alternate positive/

Table 1. Main Ion Species Observed during LC-ESI-MS Analysis of the Flavanoids from WS-2 and *tt7 A. thaliana* Seed Extracts (R, Rhamnosyl; G, Glucosyl; K, Kaempferol; Q, Quercetin)

compound	t_R (min)	MH^+ (*m/z*)	fragments and adduct ions* (ES^+) (*m/z*)	$[M-H]^-$ (*m/z*)	observed in
1 K-R-G-3-R-7	9.3	741 (55)	595 (100), 433 (56), 287 (3), 763* (24)	739 (100)	*tt7*
2 K-R-3-G-7	12.4	595 (100)	449 (45), 287 (3), 617* (19)	593 (100)	*tt7*
3 Q-G-3-R-7	14.3	611 (100)	449 (29), 303 (1), 633* (4)	609 (100)	WS-2
4 K-G-3-R-7	16.3	595 (100)	433 (28), 287 (2), 617* (31)	593 (100)	*tt7* and WS-2
5 Q-R-3-R-7	16.4	595 (100)	449 (95), 303 (6), 617* (50), 633* (3)	593 (100)	WS-2
6 K-R-3-R-7	17.9	579 (72)	433 (100), 287 (7), 601* (17)	577 (100)	*tt7* and WS-2
7 K-G-3	20.0	449 (100)	287 (35), 471* (66)	447 (100)	*tt7*
8 Q-R-3	20.1	449 (27)	303 (100), 471* (16)	447 (100)	WS-2
9 K-R-3	22.0	433 (28)	287 (100), 455* (21)	431 (100)	*tt7* and WS-2

negative ionization mode during recording was preferred to ensure the assignment of the molecular weights. **Table 1** shows the values obtained for each of the detected compounds (numbered from **1** to **9** following their retention times). In the positive mode, the $[M + H]^+$ ions are observed together with cationized adducts ($[M + Na]^+$, $[M + K]^+$) and fragment ions even at low cone voltage (25 V). Most of the fragments issuing from the $[M + H]^+$ ion seem to result from the loss(es) of 162 and/or 146 u likely corresponding to sugar residues (hexose and desoxyhexose, respectively). The negative ion MS spectra exhibit the unique $[M - H]^-$ ion under the same conditions. Further structural features were determined using tandem mass spectrometry. Second-order spectra were obtained by CID of the various molecular species and eventually fragments generated under appropriate optimized source conditions.

Monoglycosides. Monoglycoside **8** appeared as the major flavonoid of Ws-2 seed extracts. The CID spectrum of the MH^+ ion *m/z* 449 gave rise to a product ion at *m/z* 303 (Y_0^+) that corresponded to the elimination of a rhamnose residue (146 u). This spectrum also exhibited complementary low mass ions at *m/z* 147 and 129 interpreted as the B_1^+ and $[B_1 - H_2O]^+$ product ions, respectively. When a higher cone voltage is used, the CID of the *m/z* 303 "in source" product ion can also be recorded to reveal the nature of the aglycone. The pseudo-third-order product ions thus obtained at *m/z* 165 ($^{0,2}A^+$), 153 ($^{1,3}A^+$), and 137

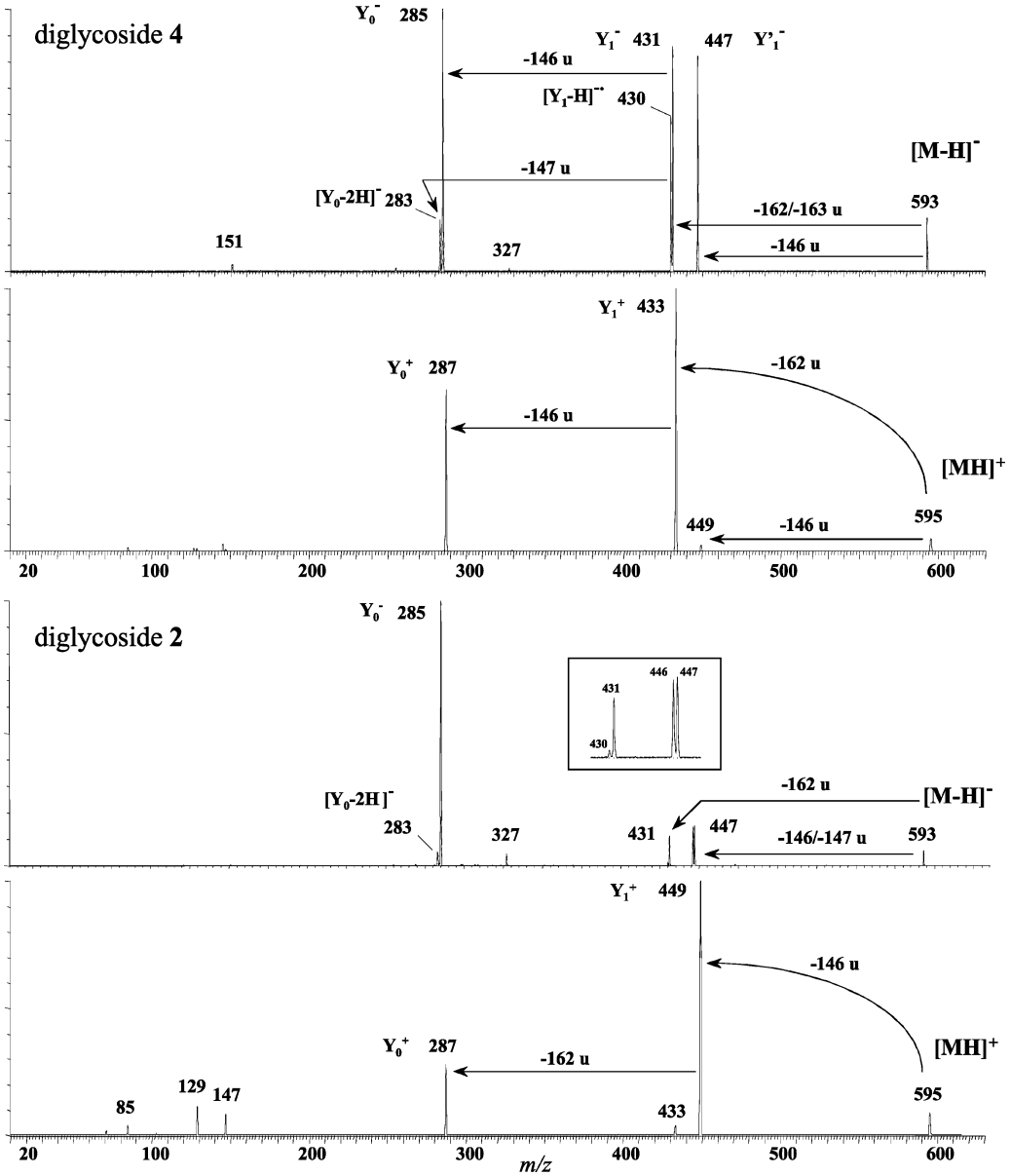


Figure 3. CID spectra of the $[M - H]^-$ and MH^+ ions of the isomeric kaempferol diglycosides **2** and **4** at 25 eV (negative mode) and 15 eV (positive mode) collision energy.

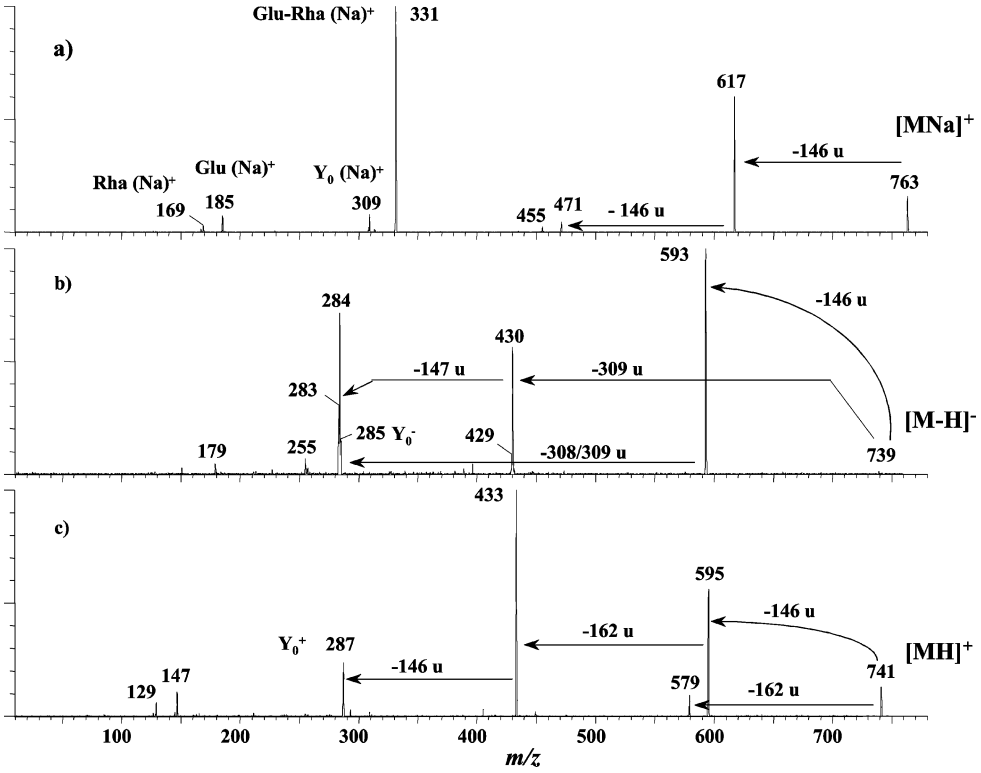
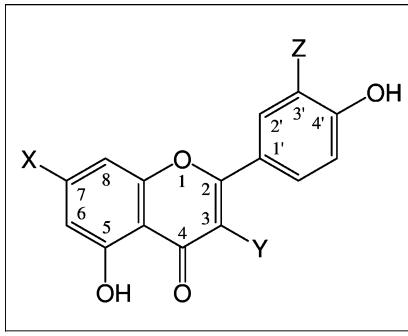


Figure 4. CID spectra of the $[M - H]^-$, MH^+, and $[MNa]^+$ ions of the kaempferol triglycoside **1** at 35 eV collision energy.



Compound	X	Y	Z
1	α-rha	α-rha(1→2)β-glc-	H
2	β-glc	α-rha	H
3	α-rha	β-glc	OH
4	α-rha	β-glc	H
5	α-rha	α-rha	OH
6	α-rha	α-rha	H
7	OH	β-glc	H
8	OH	α-rha	OH
9	OH	α-rha	H

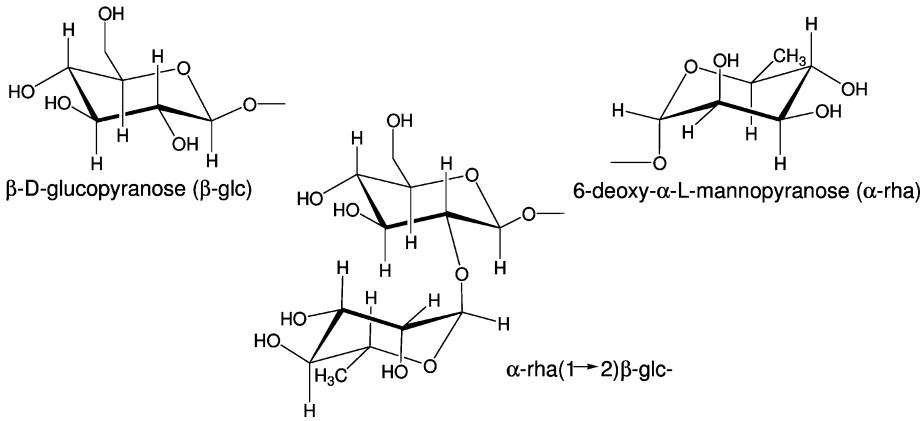


Figure 5. Flavonoid structures identified in the Ws-2 and *tt7 A. thaliana* seed extracts.

($^{0,2}B^+$) and the overall pattern were typical of the quercetin skeleton. These data and the LC-MS retention time were further found to be identical to those of the standard, quercetin 3-*O*-α-L-rhamnopyranoside. The structures of the other monoglycosides were similarly investigated. The CID spectra of the MH^+ ions *m*/*z* 433 (from **9** present in both the wild type and the *tt*-7 mutant) and 449 (issuing from **7** present only in the *tt*-7 mutant) yield the same major product ion *m*/*z* 287 (Y_0^+) by loss of a rhamnose residue or a hexose residue (glucose or galactose), respectively. When produced in source, this product ion was identified in both cases as a protonated kaempferol through the diagnostic pseudo-third-order ions *m*/*z* 165 ($^{0,2}A^+$), 153 ($^{1,3}A^+$), and 121 ($^{0,2}B^+$). Typical low mass ions were also found after the dissociation of the MH^+ ions, confirming the assignment of the rhamnosyl (*m*/*z* 147, 129) and hexosyl (*m*/*z* 145, 127 originated from the undetected *m*/*z* 163 ion by losses of H_2O)

Table 2. Chemical Shifts (δ) and Coupling Constants (J in Hertz) of Compounds **3** and **6** and HMBC Correlations, Important for Structure Elucidation

	compound **3**			compound **6**		
position	^{1}H ($^3J_{HH}$)	^{13}C ($^1J_{CH}$)	HMBC	^{1}H ($^3J_{HH}$)	^{13}C ($^1J_{CH}$)	HMBC
	Flavonoid Nucleus: Quercetin			Flavonoid Nucleus: Kaempferol		
2		159.4			159.8	
3		135.5			136.4	
5		162.9			162.8	
6	6.47 (2.2)	100.4	6 → 5, 7	6.48 (2.2)	100.4	6 → 5, 7
7		163.3			163.3	
8	6.76 (2.2)	95.4	8 → 7, 9	6.75 (2.2)	95.5	8 → 7, 9
9		157.9			157.9	
10		107.1			107.4	
1′		122.7			122.2	
2′	7.72 (2.3)	117.4	2′ → 3′, 4′, 6′, 2	7.81 (8.7)	131.8	2′ → 4′, 2
3′		145.7		6.94 (8.7)	116.4	3 → 4′, 1′
4′		149.9			161.8	
5′	6.88 (8.5)	115.9	5′ → 1′, 3′, 4′	6.94 (8.7)	116.4	5′ → 4′, 1′
6′	7.62 (2.3, 8.5)	123.2	6′ → 2′, 4′, 2	7.81 (8.7)	131.8	6′ → 4′, 2
	X: 6-Deoxy-α-mannose			X: 6-Deoxy-α-mannose		
X1	5.57 (1.6)	99.8 (174.3)	X1 → 7	5.56 (1.6)	99.8 (171.6)	X1 → 7
X2	4.02 (1.6, 3.5)	71.6		4.01 (1.6, 3.5)	71.6	
X3	3.83 (3.5, 9.5)	71.9		3.83 (3.5, 9.4)	71.9	
X4	3.48 (9.5)	73.4		3.47 (9.4)	73.5	
X5	3.59 (9.5, 6.2)	71.2		3.60 (9.5, 6.3)	71.2	
X6	1.25 (6.2)	17.9		1.26 (6.3)	17.9	
	Y: β-Glucose			Y: 6-Deoxy-α-mannose		
Y1	5.33 (7.7)	103.8 (161.4)	Y1 → 3	5.40 (1.6)	103.4 (174.3)	Y1 → 3
Y2	3.49 (7.7, 9.2)	75.6		4.22 (1.6, 3.3)	71.8	
Y3	3.43 (9.0)	77.9		3.71 (3.3, 9.3)	71.9	
Y4	3.34 (9.4)	71.2		3.33 (9.3)	73.0	
Y5	3.23 (9.8, 2.3, 5.5)	78.3		3.37 (9.5, 6.0)	71.3	
Y6	3.72, 3.56 (2.3, 5.5, 11.9)	62.5		0.94 (6.0)	17.4	

Table 3. Chemical Shifts (δ) and Coupling Constants (J in Hertz) of Compound **1** and HMBC Correlations, Important for Structure Elucidation

	compound **1**		
position	^{1}H ($^3J_{HH}$)	^{13}C	HMBC
	Flavonoid Nucleus: Kaempferol		
2		159.0	
3		134.6	
5		162.7	
6	6.45 (2.1)	100.4	6 → 5, 7
7		163.2	
8	6.76 (2.1)	95.4	8 → 7, 9
9		157.8	
10		107.5	
1′		122.8	
2′	8.08 (8.8)	132.2	2′ → 4′, 2
3′	6.90 (8.8)	116.1	3′ → 4′, 1′
4′		161.5	
5′	6.90 (8.8)	116.1	5′ → 4′, 1′
6′	8.08 (8.8)	132.2	6′ → 4′, 2
	X: 6-Deoxy-α-mannose		
X1	5.57 (1.5)	99.8	X1 → 7
X2	4.02 (1.5, 3.5)	71.6	
X3	3.82 (3.5, ∼9)	72.1	
X4	3.48 (∼9, ∼6.4)	73.6	
X5	3.59 (6.2)	71.2	
X6	1.24 (6.2)	18.1	
	Y: β-Glucose		
Y1	5.77 (7.8)	100.2	Y1 → 3
Y2	3.63 (7.8, 9.2)	80.0	Y2 → Y′1
Y3	3.56 (9.0, ∼7.7)	78.9	
Y4	3.28 (∼7.7, 9.8)	71.8	
Y5	3.23 (9.8, 2.3, 5.9)	78.5	
Y6	3.74, 3.50 (2.3, 5.9)	62.6	
	Y′: 6-Deoxy-α-mannose		
Y′1	5.24 (1.5)	102.6	Y′1 → Y2
Y′2	4.00 (1.5, 3.3)	72.4	
Y′3	3.77 (3.3, ∼10)	72.3	
Y′4	3.34 (∼10, 6.2)	73.9	
Y′5	4.03 (6.2)	69.8	
Y′6	0.95 (6.2)	17.5	

moieties. Compound **7** was compared to a standard of kaempferol 3-*O*-β-D-glucopyranoside and found to be identical (positive and negative CID spectra examined under a wide range of collision energies and retention time).

Diglycosides. The CID spectrum, of the MH^+ ion *m*/*z* 611 from compound **3** (Ws-2), showed two main product ions indicating two successive losses of sugar residues. The first loss corresponded to a hexose residue, likely a glucose (162 u), yielding Y_1^+ at *m*/*z* 449. This fragmentation was followed by the elimination of a rhamnose residue (146 u), giving the *m*/*z* 303 ion (Y_0^+), itself assigned as a protonated quercetin through its specific dissociation. This behavior is clearly related to an O-di (or di-O-) glycosylated structure versus C-glycosylation (*15*), but the question of whether the two sugars are linked (disaccharide) or not could not be solved at this stage. In this regard, the CID spectrum of the $[M - H]^-$ anion *m*/*z* 609 was more explicit. This spectrum showed two possible losses, hexose or rhamnose, yielding the product ions *m*/*z* 447 (Y_1^-) and 463 ($Y'_1{}^-$), which after the subsequent loss of the second sugar residue gave ion *m*/*z* 301 (Y_0^-). The initial competition between both types of elimination clearly relates to distinct attachment positions of the sugars to the aglycone, whereas a disaccharide would have led solely to the elimination of the global moiety in these conditions (*27*; unpublished results). A second pathway is observed particularly at higher collision energies leading to the elimination of the radical $C_6H_{11}O_5°$ (163 u) typical of the hexose and followed by that of the rhamnose also as a radical (147 u). The regiospecific obtention of such ions (*m*/*z* 446 and 299) is in favor to the location of the hexose at C-3 (*27*). Examined similarly, the diglycoside **6** (Ws-2 and *tt-7*) was found to contain two rhamnose residues and a kaempferol skeleton under the positive mode. Dissociation of the $[M - H]^-$ anion *m*/*z* 577 leading sequentially to the product ions *m*/*z* 431 and 285 (two losses of desoxyhexose molecule) on one side and to ions *m*/*z* 430 and 283 on the other (radical losses of $C_6H_{11}O_4°$) again speak for a di-O-glycosylated structure, with one rhamnosyl group attached at C-3. The same behavior observed for compound **5** (Ws-2) under both MS-MS modes indicated a quercetin analogue of **6** (di-*O*-rhamnosyl quercetin). Diglycoside **4** was coeluting in the Ws-2 extract with **5**. Because **4** was less abundant in Ws-2, it was easier to study it in the *tt-7* mutant, where the quercetin derivative was absent. Interestingly, the *tt7* mutant produced an additional diglycoside **2** that is an isomeric form of **4**, the behavior of which under MS-MS was to some extent opposite. Thus, as shown in **Figure 3**, the CID spectrum of the MH^+ ion *m*/*z* 595 from compound **4** produced successive losses of hexose (or trace of rhamnose) and then of rhamnose (or trace of hexose) to give a Y_0^+ ion at *m*/*z* 287 (protonated kaempferol as demonstrated by pseudo-third-order spectra). The reverse situation was observed for **2**, although resulting in the same final product ion *m*/*z* 287. In the negative mode, the decomposition of the $[M - H]^-$ ion *m*/*z* 593 provided in both cases two equally competitive pathways leading either to eliminate first a rhamnose (146 u) or a hexose (162) residue (*m*/*z* 447 and 431, respectively) and second the remaining sugar to give ion Y_0^- *m*/*z* 285. The main difference was the occurrence of a third pathway favored at high collision energy and corresponding to the elimination of either a hexose (**4**) or a rhamnose (**2**) radical residue, yielding product ions at *m*/*z* 430 or 446, respectively, and finally the $[Y_0 - 2H]^-$ ion *m*/*z* 283 after elimination of a second radical issuing from the remaining sugar. All of these data support di-O-glycosylated kaempferol

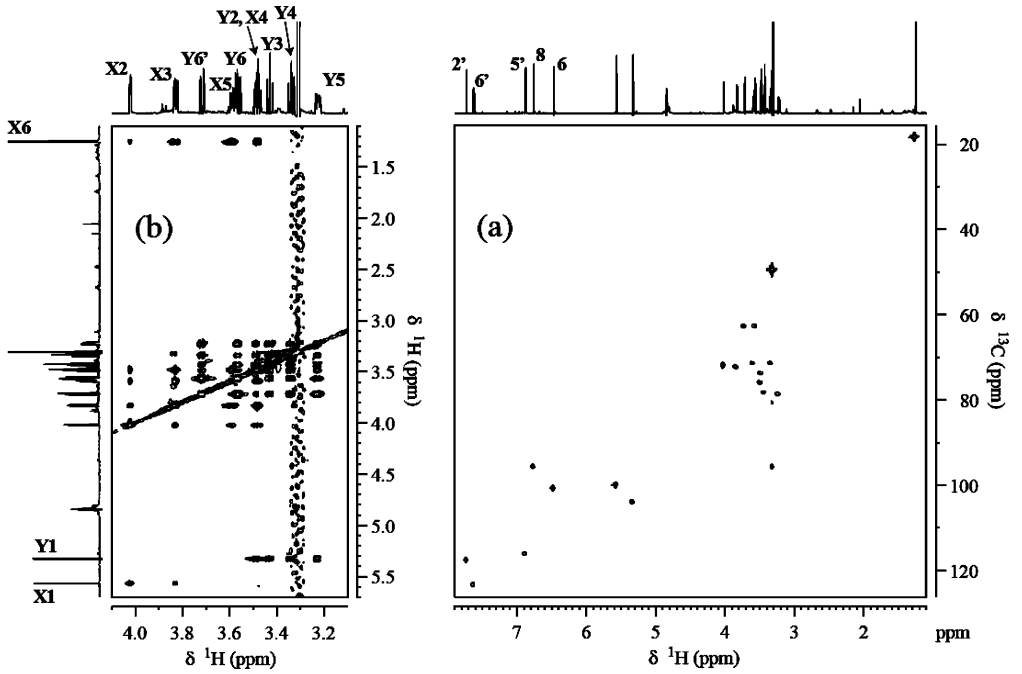


Figure 6. ^{1}H–^{13}C HSQC spectrum (**a**) and region of the ^{1}H–^{1}H TOCSY map showing sugar proton correlations (**b**) of compound **3**. Spectra were recorded with four and eight scans per transient, respectively.

structures with a hexosyl at C-3 for **4** and a rhamnosyl in the case of **2**. The second position of the glycosylation might be assumed at C-7 as this situation was in agreement with the much contrasted behavior of the two isomeric diglycosides under both the positive and negative CID modes of decomposition of the molecular species. This assignment would be also consistent with the structures of the 3,7-di-O-glycosylated flavonols described in the leaves of *A. thaliana* (*21*).

Triglycoside. According to the CID spectrum of the MH^+ ion *m*/*z* 741 (**Figure 4**), compound **1** would be a triglycosylated kaempferol containing two rhamnose and one hexose (glucose or galactose) unit. A rhamnose residue (146 u) is eliminated first, giving the product ion *m*/*z* 595, preferably to the elimination of the hexose unit (162 u) (yielding the minor ion at *m*/*z* 579). The major fragment is at *m*/*z* 433, resulting mainly from the loss of the hexose residue from *m*/*z* 595. Ion *m*/*z* 433 then decomposes into the product ion *m*/*z* 287, corresponding to the protonated aglycone (kaempferol). Low mass ions corresponding to rhamnose are also detected (*m*/*z* 147 and 129). To reduce the number of possibilities for locating the sugars, a CID experiment was conducted on the $[M + Na]^+$ adduct *m*/*z* 763 produced in source. Interestingly, this spectrum exhibited a product ion *m*/*z* 331 corresponding to the cationized diglycosyl species $[Glu\text{-}Rha]Na^+$ besides ion *m*/*z* 617 (regiospecific loss of rhamnose from the parent ion). The CID decomposition of the $[M - H]^-$ ion *m*/*z* 739 could be interpreted as resulting from the competition between two processes: elimination of an isolated rhamnose residue of 146 u to produce the major ion at *m*/*z* 593 and loss of a diglycosyl radical (309 u) yielding the radical anion *m*/*z* 430. These ions would give subsequently ions *m*/*z* 283−285 by loss of the complementary osidic group under either a molecular or radical form. This interpretation was confirmed by the pseudo-third-order CID spectrum of ion *m*/*z* 593, yielding directly ions *m*/*z* 284 and 285 by loss of 309 and 308 units. These data lead us to conclude that there is a diglycosylation by a rhamnosylhexose unit at C-3 (the most susceptible to produce the homolytic cleavage of the glycosidic bonding onto the aglycone). The second rhamnosyl group might be attached at C-7.

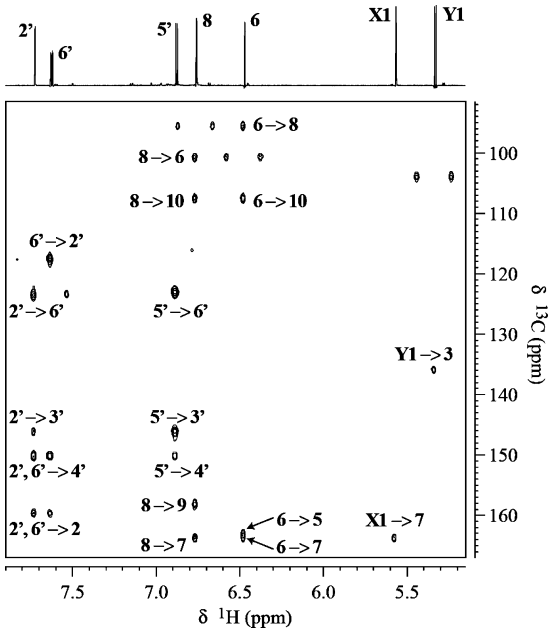


Figure 7. Region of the ^{1}H–^{13}C HMBC map used for positioning the sugar units of compound **3** to the flavonoid backbone (32 scans per transient).

Determination of the Nature and Position of the Sugar Linkages Using NMR Spectroscopy. The chemical structures of the flavonoid glycosides from *A. thaliana* seeds, compatible to their NMR spectra, are presented on **Figure 5**. The analysis of the NMR data is presented below for the cases of compounds **3** and **6** (**Table 2**) as well as compound **1** (**Table 3**). The same procedure was used for the structural investigation of the other analogues. During the initial steps of the assignment, the flavonoid part and the sugar moieties were treated separately. Extraction of the chemical shift values for the aromatic H−C pairs from the HSQC spectrum (**Figure 6a**) and their comparison to the predicted values from the ACD package (ACD Labs) as well as information from proton ^{3}J couplings resulted in the

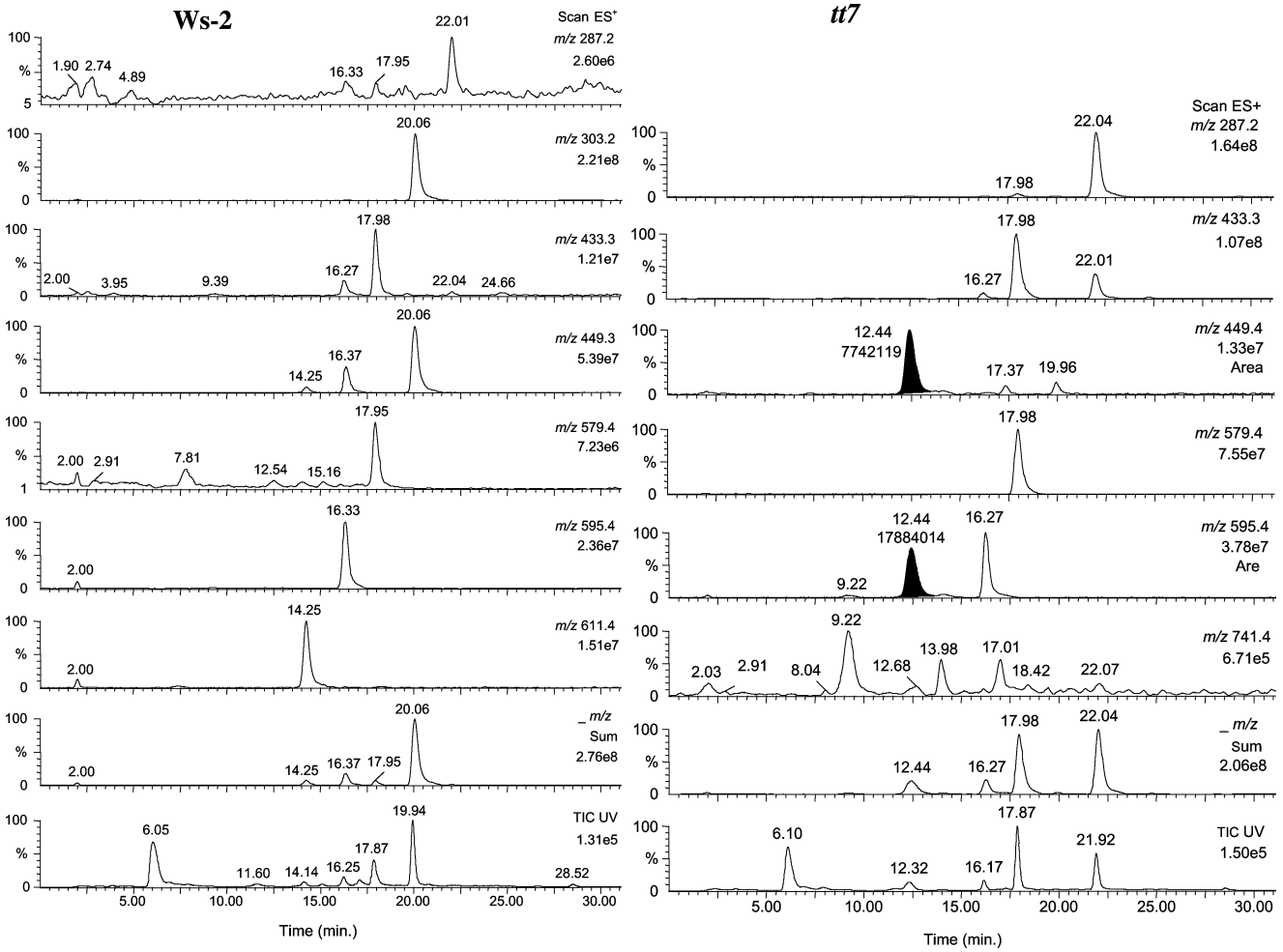


Figure 8. ESI-LC-MS profiles from (left) Ws-2 and (right) *tt7* mutant *A. thaliana* seeds: specific ion currents for MH^+ and first-generation fragment ions, sum of current ions, and UV trace (280 nm) of the major flavonoids.

attribution of the aromatic protons and their carbons. Subsequently, the HMBC experiment provided ^3J connectivities of these protons with most of the quaternary carbons, as demonstrated in **Figure 7**. All of these data confirmed that the aglycones of compounds **3** and **6** were quercetin and kaempferol, respectively, the latter presenting a simpler NMR resonance pattern for ring C due to symmetry.

Dealing with the sugars, COSY and TOCSY data allowed spin-system identification using the well-resolved anomeric protons as starting point, and this is shown in **Figure 6b** for the case of compound **3**. Determination of proton ^3J coupling constants revealed the relative configuration of vicinal protons: small couplings (<2 Hz) are indicative of equatorial−equatorial disposition, intermediate couplings (3−4 Hz) refer to equatorial−axial disposition, and large couplings (>7 Hz) correspond to an axial−axial one. Therefore, on the basis of the data of **Table 2**, moiety X was identified as 6-deoxy-α-mannose (α-rhamnose) and Y as *β*-glucose for compound **3**, whereas compound **6** contained two 6-deoxy-α-mannose residues. This was further confirmed by the determination of the $^1J_{C1-H1}$ coupling, which is 168−174 Hz for α-pyranosyl residues and 160−167 Hz for *β*-pyranosyl residues (*28*). The rather high chemical shift value of the *β*-glucose anomeric proton of compound **3** can be explained by the deshielding effect of nearby aromatic ring and double bonds.

Concerning the linkage between the flavonoid nuclei and the sugar moieties, the relevant information can be found in the HMBC spectra (**Figure 7**). The existence of cross-peaks relating pyranosyl anomeric protons to carbons of the flavonoid rings A or B (^3J long-range coupling H−C−O−C) allowed us to unambiguously establish the connectivity between sugar and aglycone parts.

The NMR study of compound **1** was the most difficult one due to the small quantity of the product available and the presence of interfering compounds in the sample, at almost similar concentrations. As previously discussed, kaempferol was identified as the flavonoid part, and three pyranoses were found to be linked to it: one *β*-glucose and two 6-deoxy-α-mannoses. Due to the extensive overlapping, a DQF-COSY experiment was used to derive $^1H-^1H$ coupling constants that could not be measured directly on the 1D spectrum. The HMBC spectrum exhibited ^3J correlations between the anomeric protons of the glucose (Y) and of one rhamnose moiety (X) with the carbons 3 and 7 of kaempferol, respectively (**Table 3**). Searching the position of the second rhamnose (Y′), we identified long-range couplings between its anomeric proton and the glucose C2 carbon and the inverse, between the glucose H2 proton and the rhamnose anomeric carbon (*29*). Moreover, inspection of the NOESY spectrum (mixing time of 400 ms) revealed a through-space proximity of rhamnose H1 and glucose H2 protons, compatible with the proposed structure of the disaccharide shown in **Figure 5**. Indeed, the distance between these two protons was found to be ≈2.3−2.4 Å in low-energy models of the molecule.

Comparative LC-MS Profiles from Ws-2 and the *tt7* Mutant *A. thaliana* Seeds. The total MS^+ ion chromatogram

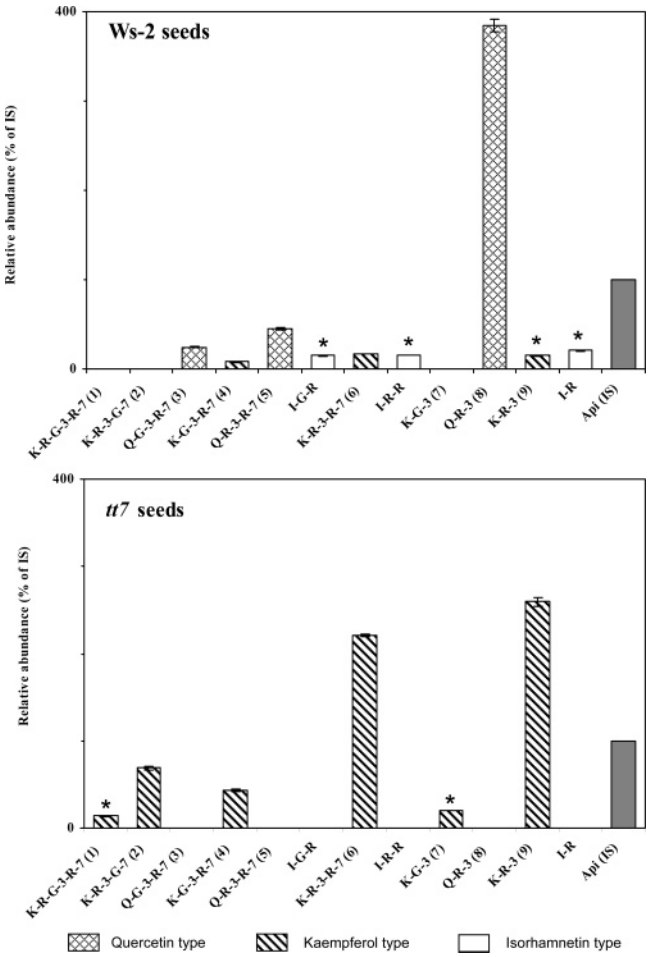


Figure 9. Comparative flavonoid compositions (histograms) of (top) Ws-2 and (bottom) *tt7* seed extracts including minor partially identified isorhamnetins, i.e., 3′-*O*-methylquercetins (*24*). * indicates a relative abundance multiplied by 5.

did not resolve distinct peaks. Instead, specific ion currents corresponding to MH^+ ions and their in-source fragments were able to exhibit characteristic flavonoid signals. This means of detection for the flavonoid metabolites was confirmed by the UV absorbance at λ 280 nm recorded in parallel. In general, two ion currents (MH^+ and the first fragment issuing from the parent ion) were found to be necessary to assess the specific detection of the flavonoid constituents. **Figure 8** shows the ion current chromatograms used specifically to detect the flavonoids contained in Ws-2 and the *tt7* mutant *A. thaliana* seeds. The *tt7* extract revealed specific flavonoid components, for example, the kaempferol diglycoside **2** appearing at 12.44 min (see ion currents at *m*/*z* 595 and 449, **Figure 8B**, right panel) and a series of triglycosides responding to ion *m*/*z* 741 (MH^+), particularly at 9.22 min (compound **1**). The search for glycosylated quercetins failed, indicating the complete absence of such constituents in the mutant. **Figure 9** summarizes the flavonoid profiles obtained, respectively, for the wild type and the mutant for a more convenient comparison. Each bar represents a given compound. Intensities corresponding to the sum of the areas in arbitrary units of the two ion signals chosen for specific detection (see example in **Figure 8**, right panel) are finally expressed as a percent of the MH^+ response of the internal standard (apigenin). Although proportions are uncorrected (exact proportions would be accessible using calibration curves of all compounds), these profiles (means of three replicates from extraction to analysis, RSD $\leq$ 7%) can be considered as representative because they remain unchanged in a certain range of concentrations (0.15−2.5 mg of seed/mL extract).

Being used complementarily, LC-ESI-MS and NMR studies were able to provide precise structures of the main flavonoids present in seeds of *A. thaliana*. The Ws-2 accession contains quercetin 3-*O*-α-rhamnopyranoside (**8**), kaempferol 3-*O*-α-rhamnopyranoside (**9**), quercetin and kaempferol 3-*O*-*β*-glucopyranoside-7-*O*-α-rhamnopyranoside (**3** and **4**), and quercetin and kaempferol 3,7-di-*O*-α-rhamnopyranoside (**5** and **6**). Investigated in parallel, the *tt7* mutant was found to produce, as expected, only kaempferol glycosides, but three additional compounds absent in Ws-2 seeds were detected, namely, kaempferol 3-*O*-*β*-glucopyranoside (**7**), kaempferol 3-*O*-α-rhamnopyranoside-7-*O*-*β*-glucopyranoside (**2**), and kaempferol 3-*O*-*β*-[α-rhamnopyranosyl(1→2)-glucopyranoside]-7-*O*-α-rhamnopyranoside (**1**). The triglycoside **1** was reported first in the leaves of *Sedum telephium* ssp. *maximum (*Crassulaceae) (*30*) and subsequently detected in *Arabidopsis* leaves (*22*). Another

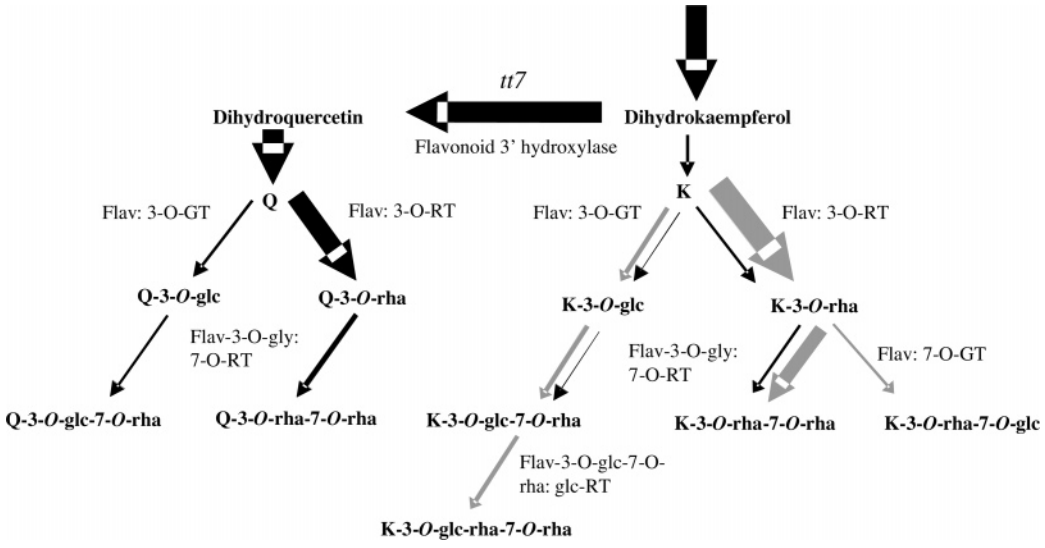


Figure 10. Proposed biosynthetic scheme based on structures of flavonol glycosides identified in Ws-2 accession and *tt7* mutant *A. thaliana* seed. Black arrows correspond to flavonol glycoside accumulation in Ws-2 seed and gray arrows to kaempferol glycoside accumulation in *tt7* seed. Line widths are related to product accumulation. Enzymes are annotated beside arrows. Products are in black. Flav, flavonol; GT, glucosyltransferase; RT, rhamnosyltransferase; K, kaempferol; Q, quercetin; glc, glucoside; gly, glycoside; rha, rhamnoside. Adapted from Jones et al. (*23*).

triglycoside with a gentobioside (glc 1→6 glc) rather than a neohesperoside (rha 1→2 glc) substitution was also found in *Arabidopsis* leaves (*21*).

Ws-2 seed flavonol glycosides are clearly different from those reported in the leaves by a higher ratio of quercetin/kaempferol aglycone derivatives in seeds and the specific occurrence of either monoglycosides (seeds) or triglycosides (leaves). On the other hand, the flavonoid composition of *tt7* seeds differs (**Figure 9**) from that found for the wild type (Ws-2) by the lack of quercetin derivatives and a number of criteria regarding the glycosylation process. Effectively, although rhamnosylation (vs glucosylation) on OH-3 is the preferred first step in both cases, a higher accumulation of di- and triglycosylation is observed next in *tt7* as compared to Ws-2 (**Figure 10**). The second sugar is attached at the 7-OH again through rhamnosylation preferentially, but not exclusively. Particularly remarkable is the occurrence of kaempferol-3-*O*-rhamnoside-7-*O*-glucoside **2** in *tt7* seeds, which indicates the possible 7-O-glucosylation of kaempferol 3-*O*-α-rhamnopyranoside (**9**). Although the gene coding for the flavonol-7-*O*-glucosyltransferase was identified in *Arabidopsis* leaves and flowers and shown to preferentially use in vitro quercetin-3-rhamnopyranoside (*23*), the occurrence of diglycoside **2** suggests that kaempferol may also be an in vivo substrate for this enzyme. Interestingly, a possible quercetin analogue was shown to temporarily accumulate during Ws-2 seed development (*24*). The flavonoid-3′-hydroxylase may thus have an important role, besides other factors such as the availability and/or the affinity of a given substrate for the glycosyltransferases, to determine flavonol accumulation as revealed, first, by comparing Ws-2 versus *tt7* seeds and, second, seeds versus leaves in the wild type. The presence of compounds **1** and **2** in *tt7* seed (**Figure 10**) may result from the higher accumulation of kaempferol derivatives in the mutant [+63% compared to the total flavonol content in Ws-2 seeds (*24*)] or from the difference in the glycosyltransferase enzyme affinities for flavonols.

Like other seeds and grains such as soybean, pea, or bean, *Arabidopsis* essentially contains quercetin and kaempferol glycosides and, like almond and peanut, it also accumulates isorhamnetin derivatives (*24*). Moreover, most aglycones are glycosylated in positions C-3 and C-7 and the preferred sugar substitutions are glucose and rhamnose, even if other glycosides may be found (for instance, galactoside, rutinoside, xyloside, or acetylglucoside) in other species (*31*). The data reported here confirm that *Arabidopsis* could be a valuable model for a better understanding of the flavonol metabolism in other seeds.

ACKNOWLEDGMENT

We acknowledge M. Caboche (INRA, Versailles, France) for initiating the project, and we thank I. Debeaujon (INRA) for helpful discussions.

LITERATURE CITED

(1) Dixon, R. A. Engineering of plant natural product pathways. *Curr. Opin. Plant Biol.* **2005**, *8*, 329−336.
(2) Schijlen, E. G.; Ric de Vos, C. H.; van Tunen, A. J.; Bovy, A. G. Modification of flavonoid biosynthesis in crop plants. *Phytochemistry* **2004**, *65*, 2631−2648.
(3) Marles, M. A.; Ray, H.; Gruber, M. Y. New perspectives on proanthocyanidin biochemistry and molecular regulation. *Phytochemistry* **2003**, *64*, 367−383.
(4) Koornneef, M. Mutation affecting the testa color in *Arabidopsis*. *Arabidopsis Inf. Serv.* **1990**, *20*, 89−92.
(5) Winkel-Shirley, B. Flavonoids in seeds and grains: physiological function, agronomic importance and the genetic of biosynthesis. *Seed Sci. Res.* **1998**, *8*, 515−522.
(6) Winkel-Shirley, B. Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology and biotechnology. *Plant Physiol.* **2001**, *126*, 485−493.
(7) Schoenbohm, C.; Martens, S.; Eder, C.; Forkmann, G.; Weisshaar, B. Identification of the *Arabidopsis thaliana* flavonoid 3′-hydroxylase gene and functional expression of the encoded P450 enzyme. *Biol. Chem.* **2000**, *381*, 749−753.
(8) Peer, W. A.; Bandyopadhyay, A.; Blakeslee, J. J.; Makam, S. N.; Chen, R. J.; Masson, P. H.; Murphy, A. S. Variation in expression and protein localization of the PIN family of auxin efflux facilitator proteins in flavonoid mutants with altered auxin transport in *Arabidopsis thaliana*. *Plant Cell* **2004**, *16*, 1898−1911.
(9) Winkel-Shirley, B. Biosynthesis of flavonoids and effects of stress. *Curr. Opin. Plant Biol.* **2002**, *5*, 218−223.
(10) Lee, G. L. Condensed tannins in some forage legumes: their role in ruminant pasture bloat. *Basic Life Sci.* **1992**, *59*, 915−934.
(11) Waghorn, G. C.; McNabb, W. C. Consequences of plant phenolic compounds for productivity and health of ruminants. *Proc. Nutr. Soc.* **2003**, *62*, 383−392.
(12) Bagchi, D.; Sen, C. K.; Ray, S. D.; Das, D. K.; Bagchi, M.; Preuss, H. G.; Vinson, J. A. Molecular mechanisms of cardioprotection by a novel grape seed proanthocyanidin extract. *Mutat. Res.* **2003**, *523*−*524*, 87−97.
(13) Serafini, M.; Bugianesi, R.; Maiani, G.; Valtuena, S.; De Santis, S.; Crozier, A. Plasma antioxidants from chocolate. *Nature* **2003**, *424*, 1013.
(14) Park, O. J.; Surh, Y. J. Chemopreventive potential of epigallocatechin gallate and genistein: evidence from epidemiological and laboratory studies. *Toxicol. Lett.* **2004**, *150*, 43−56.
(15) Harborne, J.; Baxter, H. *A Handbook of the Natural Flavonoids*; Wiley: New York, 1999.
(16) Williams, C. A.; Grayer, R. J. Anthocyanins and other flavonoids. *Nat. Prod. Rep.* **2004**, *21*, 539−573.
(17) Stobiecki, M. Application of mass spectrometry for identification and structural studies of flavonoid glycosides. *Phytochemistry* **2000**, *54*, 237−256.
(18) Cuyckens, F.; Rozenberg, R.; de Hoffmann, E.; Claeys, M. Structure characterization of flavonoid *O*-diglycosides by positive and negative nano-electrospray ionization ion trap mass spectrometry. *J. Mass Spectrom.* **2001**, *36*, 1203−1210.
(19) Cuyckens, F.; Claeys, M. Mass spectrometry in the structural analysis of flavonoids. *J. Mass Spectrom.* **2004**, *39*, 1−15.
(20) Graham, T. L. Flavonoid and flavonol glycosides metabolism in *Arabidopsis*. *Plant Physiol. Biochem.* **1998**, *36*, 135−144.
(21) Veit, M.; Pauli, G. F. Major flavonoids from *Arabidopsis thaliana* leaves. *J. Nat. Prod.* **1999**, *62*, 1301−1303.
(22) Bloor, S. J.; Abrahams, S. The structure of the major anthocyanin in *Arabidopsis thaliana*. *Phytochemistry* **2002**, *59*, 343−346.
(23) Jones, P.; Messner, B.; Nakajima, J. I.; Schaffner, A. R.; Saito, K. UGT73C6 and UGT78D1, glycosyltransferases involved in flavonol glycoside biosynthesis in *Arabidopsis thaliana*. *J. Biol. Chem.* **2003**, *278*, 43910−43918.
(24) Routaboul, J. M.; Kerhoas, L.; Debeaujon, I.; Pourcel, L.; Caboche, M.; Einhorn, J.; Lepiniec, L. Flavonoid diversity and biosynthesis in seed of *Arabidopsis thaliana*. *Planta* **2006**, *224*, 96−107.
(25) Domon, B.; Costello, C. E. A systematic nomenclature for carbohydrate fragmentations in FAB MS/MS spectra of glycoconjugates. *Glycoconjugate J.* **1988**, *5*, 397−409.
(26) Ma, Y. L.; Li, Q. M.; Van den Heuvel, H.; Claeys, M. Characterization of flavone and flavonol aglycones by collision-induced dissociation tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* **1997**, *11*, 1357−1164.

(27) Hvattum, E.; Ekeberg, D. Study of the collision-induced radical cleavage of flavonoid glycosides using negative electrospray ionization tandem quadrupole mass spectrometry. *J. Mass Spectrom.* **2003**, *38*, 43−49.

(28) Bock, K.; Pedersen, C. A study of CH coupling constants in hexopyranoses. *J. Chem. Soc., Perkin Trans. 2* **1974**, 293−297.

(29) Duus, J. O.; Gotfredsen, C. H.; Bock, K. Carbohydrate structural determination by NMR spectroscopy: modern methods and limitations. *Chem. Rev.* **2000**, *100*, 4589−4614.

(30) Mulinacci, N.; Vincieri, F. F.; Baldi, A.; Bambagiotti-Alberti, M.; Sendl, A.; Wagner, H. Flavonol glycosides from *Sedum telephium* subspecies *maximum* leaves. *Phytochemistry* **1995**, *38*, 531−33.

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

Received for review April 13, 2006. Revised manuscript received July 3, 2006. Accepted July 6, 2006. This work was supported by Genoplante (NO2001 00045).

JF061043N