Major Flavonoids from *Arabidopsis thaliana* Leaves†

Markus Veit‡ and Guido F. Pauli*,§

Department of Pharmaceutical Biology, University of Würzburg, Julius-von-Sachs-Platz 2, D-97082 Würzburg, Germany, and Institute of Pharmaceutical Biology and Phytochemistry, Westfälische Wilhelms-University, Hittorfstrasse 56, D-48149 Münster, Germany

Received March 3, 1999

The three major flavonoids isolated from *Arabidopsis thaliana* plants grown in the greenhouse were identified by means of spectroscopic analysis (UV, NMR, MS) and chiral capillary zone electrophoresis as the novel kaempferol 3-*O*-*β*-[*β*-D-glucopyranosyl(1→6)D-glucopyranoside]-7-*O*-α-L-rhamnopyranoside (**1**), kaempferol 3-*O*-*β*-D-glucopyranoside-7-*O*-α-L-rhamnopyranoside (**2**), and kaempferol 3-*O*-α-L-rhamnopyranoside-7-*O*-α-L-rhamnopyranoside (**3**). Comprehensive NMR studies including selective 1D and gradient-enhanced 2D techniques were applied in order to achieve full signal assignment and definitive proof of linkage for compound **1**.

More than 10 000 papers have been published on various aspects of the biology, chemistry, and genetics of *Arabidopsis thaliana* L. (Brassicaceae) in the last three decades. About 200 of these deal with the molecular biology, biosynthesis, and ecology of the flavonoids. Surprisingly, the structures of the phenolics which accumulate in *A. thaliana* have not been studied in detail. It was only very recently when the leaf flavonoids were reported using HPLC and subsequent hydrolysis in addition to UV, MS, and NMR spectroscopy. While the results revealed glucose and rhamnose moieties that were attached to kaempferol and quercetin, respectively, no evidence was provided for essential structural features such as the inter-glycosidic linkages and the linkages of the sugars to the aglycon as well as the absolute chirality of the sugars.[1] However, exact identification of the glycosides is a prerequisite for further research on the biochemistry and ecology of *Arabidopsis* leaf flavonoids, which focuses mainly on the genetics, enzymology, and regulation of flavonoid biosynthesis and their role in plant protection against UV light.[1−3] These experiments are intended to provide major insights into the chemistry and biology of plants in general.

When grown in the greenhouse or under artificial conditions without supplemental UV light, *A. thaliana* accumulates only kaempferol glycosides.[4] Because the corresponding quercetin metabolites can be detected only after UV exposure of the plants[4] and represent critical pairs of preparative separation, we have focused on specimens grown in the greenhouse and now report the isolation and identification of the main leaf flavonoids from UV unexposed *A. thaliana* L. (ws-2).

Compounds **1**−**3** were isolated by repeated column chromatography on polyamide and Sephadex LH-20 using the methanol extract of freeze-dried and ground *A. thaliana* leaves. The structure elucidation of compound **1** and the identification of the already known **2** and **3** was carried out by UV, MS, and extensive 1D and 2D ^{1}H and ^{13}C NMR studies. The molecular masses resulted from the $[M - H]^-$ and $[M + Na]^+$ ions observed in the API-ESI and from HR-MALDI-TOF mass spectra and were determined as 756 (**1**), 594 (**2**), and 578 amu (**3**), respectively. Due to the fragmentation pattern in MS^n experiments, HR-MS measurements (**1**, calcd for $C_{33}H_{40}O_{20}$) and augmented by both the integration of the ^{1}H spectra and the number and multiplicity of the ^{13}C resonances, the compounds were assigned as glycosides containing two (**2** [hexose + 6-desoxyhexose] and **3** [two hexoses]) or three (**1** [two hexoses + 6-desoxyhexose]) sugars, respectively. The UV spectra supported the structure of kaempferol glycosides, and upon addition of UV shift reagents the presence of a 3,7-di-*O*-substitution pattern was indicated.[5]

Definitive evidence for the sugar linkage concerning both the flavonoid-sugar and the inter-glycosidic bondings, came from detailed 1D and 2D NMR measurements including high-field ^{1}H, ^{13}C and APT, gradient-enhanced COSY as well as gradient-selected HSQC and HMBC experiments. Selective pulse 1D TOCSY and NOE measurements were employed to establish the relative stereochemistry and linkage of the sugars, respectively. Furthermore, spectral simulation certified the analysis of the higher order spin systems in both the sugar and the aglycon moieties. As a result, fully assigned sets of proton and carbon resonances were obtained. The flavonoid skeleton gave rise to typical AA′BB′ proton signals centered around 8.0 and 6.9 ppm and two 2.1 Hz doublets for the A-ring protons H-6 and H-8. Definitive assignments of the A/C ring carbon signals was achieved through the gHSQC/gHMBC maps. The relative stereochemistry of the hexopyranoses was clearly established from the ^{1}H NMR coupling pattern as follows: *β*-glucose (glc) revealed a typical all *trans*-diaxial coupling pattern within the pyranoid ring system yielding only large coupling constants (7.7−9.5 Hz) for H-1 through H-5 plus additional couplings to the hydroxymethylene protons H-6A/B (2.4 and 5.5 Hz). The relative stereochemistry of α-rhamnose (rha) representing the 6-desoxy derivative of the C-2 epimer of glucose, i.e., mannose, was deduced from the small couplings (1.9 and 3.5 Hz) between H-2 and its neighbors. It must be mentioned that in some cases, e.g., the anomeric protons of 7-*O*-rha and 3-*O*-glc as well as the protons H-4/5/6 of 7-*O*-rha and H-2/3 of 3-*O*-glc, the presence of higher order spin systems requires labeling of the signals as multiplets and analysis of the active couplings using spectral simulation. Further evidence for these observations came from 1D selective TOCSY experiments allowing the acquisition of undivided subspectra of each particular carbohydrate moiety as a result of the presence of isolated sugar spin cages. Accordingly, upon excitation

* To whom correspondence should be addressed. Tel.: +49 251 833 3376. Fax: +49 251 833 2114. E-mail: pauli@uni-muenster.de.

† Dedicated to Professor R. Hegnauer on the occasion of his 80th birthday.

‡ Department of Pharmaceutical Biology, University of Würzburg.

§ Institute of Pharmaceutical Biology and Phytochemistry, WWU Münster.

10.1021/np990080o CCC: $18.00
Published on Web 08/14/1999

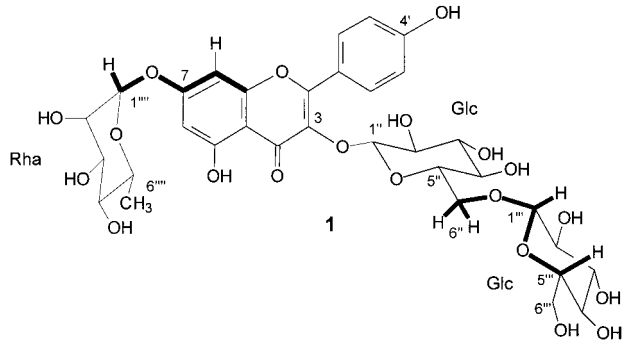

Figure 1. Structure of **1** with bold lines indicating selected, diagnostic $^1H-^{13}C$ long-range correlations observed in the HMBC spectrum. The 6‴-*O* linkage of the terminal glc moiety gives rise to five anomeric $^3J_{C,H}$ connectivities, i.e., three over the glycosidic bond and two involving a prominent H-5‴ shifted to unusually high field. The 7-*O*-L-rha partial structure was evident from a $^{2/3}J_{C,H}$ pathway connecting H-1⁗ and C-9.

of the anomeric proton, magnetization could easily be transferred throughout the whole ring system in glc due to the large *trans*-diaxial *J* values. The best excitation frequency in rha was not H-1 ($^3J_{1,2}$ = 1.9 Hz) but H-2, because the reasonable size of $^3J_{2,3}$ = 3.5 Hz allowed full magnetization transfer toward Me-6 using a long mixing time (τ_{mix} = 140 ms).

Following from the corresponding $^3J_{C\text{-}7,Hanom}$ HBMC correlations as well as nearly identical sets of ^1H and ^{13}C NMR resonances, all compounds were shown to bear a 7-*O*-α-rha moiety. In **3** an additional α-rha moiety was found, and compound **2** contained an extra *β*-glc unit, with both sugars being attached to the OH-3 group (HMBC: $^3J_{H-1'',C\text{-}3}$). The disaccharide gentiobiose was identified as the 3-*O* substituent in **1** according to the observation of significantly shifted signals of C/H-6″ and to ^3J correlations for the pairs H-5‴/C-1‴, C-1‴/H-6″A+B, as well as H-1‴/C-6″ (see Figure 1). Distinction of the two rha moieties in **3** was possible through NOEs observed between H-1‴ and H-6/8 indicating a 7-*O*-linkage in conjunction with COSY correlations.

Finally, the absolute stereochemistry of the sugar residues in **1**−**3** was determined by capillary zone electrophoresis of their *S*-(−)-1-phenylethylamine derivatives after hydrolysis of the compounds (see Experimental Section). As a result, all *β*-glc moieties were shown to belong to the D-series, while all α-rha units exhibited L-configuration. In summary, the flavonoids were characterized as the novel kaempferol 3-*O*-*β*-[*β*-D-glucopyranosyl(1→6)D-glucopyranoside]-7-*O*-α-L-rhamnopyranoside (kaempferol 3-*O*-gentiobioside-7-*O*-rhamnoside) (**1**), beside the previously known[6−8] kaempferol 3-*O*-*β*-D-glucopyranoside-7-*O*-α-L-rhamnopyranoside (**2**) and kaempferol 3-*O*-α-L-rhamnopyranoside-7-*O*-α-L-rhamnopyranoside (**3**).

Upon exposure of the plants to UV light, the corresponding OH-3′ derivatives of **1**−**3**, i.e., the quercetin glycosides could be detected by RP-18 HPLC. These six compounds typically occur in wild types of *A. thaliana* of different ecotypes.[1] They are accumulated together with minor flavonoids which contribute less than 10% to the total flavonoid content as determined by HPLC analysis calibrating with kaempferol 3-*O*-glucoside at *λ* 350 nm. The determination of the flavonoid pattern of mutants[2,9−11] is in progress.[12]

Table 1. ^1H and ^{13}C NMR Data for **1**[a]

position	^{13}C δ[b]	^1H δ[c] (multiplicity; $J_{H,H}$ in Hz)	HMBC correlations (H → C)
2	159.3		
3	135.6		
4	179.6		
5	163.3		
6	100.7	6.49 (d; 2.1)	5,7,8,10,
7	163.6		
8	95.9	6.78 (d; 2.1)	6,7,9,10
9	158.0		
10	107.5		
1′	122.5		
2′	132.5	8.14_6 (pseudo d/m; 2.1, 7.9)	4′,6′
3′	116.3	6.91_3 (pseudo d/m; 2.1, 7.9)[d]	1′,4′,5′
4′	161.8		
5′	116.3	6.91_0 (pseudo d/m; 2.1, 7.9)[d]	1′,3′,4′
6′	132.5	8.14_4 (pseudo d/m; 2.1, 7.9)[d]	2′,4′
1″	103.5	5.35 (d; 7.7)	–
2″	75.8	3.45 (pseudo dd/m; 7.7, 9.1)	3″
3″	77.9	3.42_1 (pseudo dd/m; 9.1, 9.5)	5″
4″	71.4	3.35 (dd/t; 9.5, 9.5)	3″,5″
5″	77.9	3.42_2 (ddd; 2.0, 5.5, 9.5)	
6″	69.5	3.98 [A] (dd; 2.0, 11.8)	1‴,4″
		3.64 [B] (dd; 5.5, 11.8)	1‴,5″
1‴	104.6	4.14 (d; 7.7)	6″
2‴	75.1	3.04 (dd; 7.7, 9.1)	4‴
3‴	77.8	3.13 (dd; 9.1, 9.5)	2‴,4″
4‴	71.3	3.20 (dd/t; 9.5, 9.5)	2‴,6‴
5‴	77.7	2.97 (ddd; 2.4, 5.5, 9.5)	3‴
6‴	62.5	3.97 [A] (dd; 2.0, 11.8)	
		3.56 [B] (dd; 5.5, 11.8)	5‴
1⁗	100.0	6.58 (d; 1.9)	7,2‴,3‴
2⁗	71.6	4.01 (dd; 1.9, 3.5)	
3⁗	72.0	3.82 (dd; 3.5, 9.5)	
4⁗	73.5	3.47 (pseudo t; 9.5)	2⁗,5⁗
5⁗	71.3	3.60 (pseudo dq/m; 5.9, 9.5)	
6⁗	18.2	1.26 (pseudo d/m; 5.9)	4⁗,5⁗

[a] All NMR data were recorded using CD_3OD containing 5% of DMSO-d_6; δ values in ppm. [b] Recorded at 150 MHz. [c] Recorded at 600 MHz. [d] The protons H-2′/6′ and H-3′/5′ have slightly different shift values, and the signals show significant higher order effects.

Experimental Section

General Experimental Procedures. UV spectra were taken on a Perkin-Elmer Lambda 2 spectrometer in MeOH, IR data were acquired on a Bio-Rad model FTS 135 FT-IR instrument acquiring 128 scans in a window of 4000 to 500 cm^{-1}. The optical rotation was measured with a Perkin-Elmer 241 polarimeter. API-MS were obtained on a Finnigan LCQ instrument in the direct inlet mode scanning for both negative and positive ions over a range of *m*/*z* 150 to 1700. HR-MALDI-TOF spectra were measured on TofSpec 2E (Micromass Co., U.K.) and Lazarus III DE (self-construction, Dr. H. Luftmann, Organic Chemistry Department, WWU Münster, Germany) instruments using THAP and DHB matrixes, respectively. PEG 600 was used for calibration, with 8 and 10 single scans being accumulated, respectively, and the exact mass of the molecular ions $[M + Na]^+$ was determined by RMS calculation (TofSpec 2E) or by performing Gaussian fits to each peak and calculating a linear correction factor with respect to the PEG reference (Lazarus III DE). The NMR spectra were recorded on Varian Gemini 200 (5 mm dual probe) and Unity 600 (5 mm multinuclear probe, gradient unit) instruments operating at 200/600 MHz for ^1H, and 50/150 MHz for ^{13}C, respectively. The manufacturer's software was used for APT, gradient-enhanced COSY (gCOSY), as well as the inverse-detected gradient-selected heteronuclear correlations gHMBC (8.5 Hz) and gHSQC (145 Hz). Chemical shifts (δ in ppm) were referenced to the solvent as internal standard (3.300 and 49.00 ppm, respectively), and the coupling constants (*J*) are given in Hz. The digital resolution was better than 0.2 Hz in the ^1H and 0.6 Hz in the ^{13}C domain. The 1D selective TOCSY and NOE experiments were performed at 600 MHz using eburp-1/25 selective pulses. The data were processed offline with the Nuts 2D program package.

Plant Material. Plants of *A. thaliana* L. ecotype Wassilewskija (ws-2) were grown from seeds obtained through NASC

(University of Nottingham, Nottingham, U.K.) by cultivation in the greenhouse under ambient conditions and were harvested just before the flowering stage. The leaves were immediately frozen and freeze-dried. Voucher specimens are deposited at the Department of Pharmaceutical Biology, University of Würzburg.

Extraction and Isolation. Pulverized plant material (240 g) was extracted sequentially with 80% MeOH, MeOH, and MeOH−NH_3 in a percolator. The combined extracts were concentrated under reduced pressure and fractionated by partition with EtOAc. Further purification was achieved by column chromatography on polyamide (Macherey & Nagel SC-6) using a water → MeOH gradient followed by repeated column chromatography on Sephadex LH-20 with EtOH−H_2O mixtures as eluents. Final purification was done by column chromatography on Sephadex LH-20 with MeOH−H_2O−Me_2CO (1:1:2) to afford pure **1** (9 mg), **2** (25 mg), and **3** (56 mg) in yields of 0.0233%, 0.0104%, and 0.0038% (w/w), respectively. TLC was conducted on precoated Si gel 60 F_{254} plates (Merck) using EtOAc−water−HOAc−HCOOH (100:26:11:11).

Chiral Sugar Analysis. Determination of the D/L-configuration of the sugars was performed by chiral capillary zone electrophoresis on a Beckman P/ACE 5010 instrument using an uncoated fused silica capillary (570 mm, 50 μm i.d., 30 kV, 27 °C) and a DAD UV detector (λ 200 nm) operated with System Gold software. Conditions were chosen based on a recently published method[13] which was optimized and allowed the analysis of small quantities (0.5−1.5 mg of glycoside). Upon hydrolysis with aqueous 23% TFA solution, the *S*-(−)-1-phenylethylamine derivatives of the free sugars were formed and reacted with $NaBH_3CN$. While D/L-glc and L-rha were purchased (Fluka, Aldrich), the D-rha reference was generated by hydrolyzing 1-*O*-Me-2,3-acetonido-α-D-rhamnofuranoside with aqueous 23% TFA. Using a 50 mM $Na_2B_4O_7$ buffer at pH 10.3 that contained 2.7 M MeCN, the migration times (t_m) of the D/L sugar derivatives differed significantly as follows: Δt_m(rha) 0.16 min with D-rha 9.51 min, L-rha 9.67 min; Δt_m(glc) 0.11 min with D-glc 10.80 min, L-glc 10.69 min.

Kaempferol 3-*O*-*β*-[*β*-D-glucopyranosyl(1→6)D-glucopyranoside]-7-*O*-α-L-rhamnopyranoside (1): yellow amorphous solid (9 mg); TLC R_f 0.18; $[\alpha]_D$ 20−22° (*c* 0.1, MeOH); IR (KBr) ν_{max} 3400, 2983, 2937, 1662, 1606, 1514, 1494, 1448, 1378, 1352, 1212, 1180, 1130, 1088, 1061, 1034, 974, 916, 834, 812 cm^{-1}; UV ((MeOH) λ_{max} (log ϵ) 267 (4.34), 349 (4.24) nm; + $AlCl_3$ 272, 300 (sh), 340, 381 (sh) nm; + $AlCl_3$/HCl 274, 298 (sh), 344, 387 nm; + NaOAc 269, 349 nm; + NaOAc/H_3BO_3 269, 349 nm; + NaOMe 241 (sh), 275, 391 nm; ^{1}H (600 MHz, CD_3OD + 5% DMSO-d_6) and ^{13}C NMR (125 MHz, CD_3OD + 5% DMSO-d_6) see Table 1; positive and negative ion API-MS (MeOH) *m/z* 755 $[M - H]^-$ (100) and 779 $[M + Na]^+$ (100), 1512 $[2M - H]^-$ (3), and 1536 $[2M + Na]^+$ (4), 609 [M − (rha-H_2O)-H]$^-$ (3); API-MS^2 633 [M − (rha-H_2O) + Na]$^+$ (9); API-MS^3 471 [M − (rha-H_2O) − (glc-H_2O) + Na]$^+$ (3), 347 [2(glc-H_2O) + Na]$^+$ (100), 309 [M − (rha-H_2O) − 2(glc-H_2O) + Na]$^+$ (1); (+)-HR-MALDI-TOF *m/z* 779.2000 ± 0.0062 (TofSpec 2E) and 779.2124 ± 0.0202 (Lazarus III DE) calcd for $C_{33}H_{40}O_{20}$ $[M + Na]^+$ 779.20106, Δ 1.1 and 11 mmu, respectively.

Kaempferol 3-*O*-*β*-D-glucopyranoside-7-*O*-α-L-rhamnopyranoside (2) and kaempferol 3-*O*-α-L-rhamnopyranoside-7-*O*-α-L-rhamnopyranoside (3): UV, MS, and ^{1}H and ^{13}C NMR data were in agreement with literature values,[6−8] while the interpretation of the ^{1}H *J* values and coupling pattern has to be revised in analogy to Table 1.

Acknowledgment. The authors thank M. Eckert (University of Würzburg) for providing the plant material. Furthermore, we are grateful to the following colleagues at WWU Münster: Dr. K. Bergander and Dr. K. Bergenthal for NMR support and Mr. T. Meiners for running the LCQ mass spectra. Special thanks are due to Dr. H. Luftmann for providing HR-MS support, to Dr. M. Lechtenberg and Mrs. B. Quandt for performing the capillary zone electrophoresis, and to Mr. G. Jürgenliemk for a sample of D-rhamnose.

References and Notes

(1) Graham, T. *Plant Physiol. Biochem.* **1998**, *36*, 135−144.
(2) Bharti, A.; Khurana, J. *Photochem. Photobiol.* **1997**, *65*, 765−776.
(3) Shirley, B.; Kubasek, W.; Storz, G.; Bruggemann, E.; Koornneef, M.; Ausubel, F.; Goodman, H. *Plant J.* **1995**, *8*, 869−871.
(4) Veit, M. Unpublished work.
(5) Mabry, T.; Markham, K.; Thomas, M. *The Systematic Identification of Flavonoids*; Springer-Verlag: Berlin, 1970.
(6) Sharaf, M.; ElAnsari, M. A.; Saleh, N. A. M. *Biochem. Syst. Ecol.* **1997**, *25*, 161−166.
(7) Iwashina, T.; Matsumoto, S.; Nishida, M.; Nakaike, T. *Biochem. System. Ecol.* **1995**, *23*, 283−290.
(8) Özden, S.; Dürüst, N.; Toki, K.; Saito, N.; Honda, T. *Phytochemistry* **1998**, *49*, 241−245.
(9) Landry, L.; Chapple, C.; Last, R. *Plant Physiol.* **1995**, *109*, 1159−1166.
(10) Li, J.; Ou-Lee, T.; Raba, R.; Amundson, R.; Last, R. *Plant Cell* **1993**, *5*, 171−179.
(11) Louis, R.; Buchanan, B. *Planta* **1994**, *194*, 505−509.
(12) Further details of the HPLC analysis as well as studies into the chromatographic characterization of the flavonoid pattern of *A. thaliana* wild types and mutants will be reported in due course.
(13) Noe, C.; Freissmuth, J. *J. Chromatogr. A* **1995**, *704*, 503−512.

NP990080O