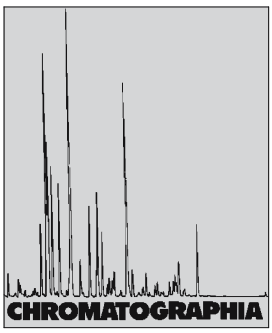
Building-Up a Comprehensive Database of Flavonoids Based on Nuclear Magnetic Resonance Data



2006, *64*, 503–508

S. Moco[1,2], Li-Hong Tseng[3], M. Spraul[3], Zheng Chen[1], J. Vervoort[1,✉]

[1] Laboratory of Biochemistry, Wageningen University, 6703 HA Wageningen, The Netherlands; E-Mail: jacques.vervoort@wur.nl
[2] Plant Research International, 6700 AA Wageningen, The Netherlands
[3] Bruker Biospin GmbH, Rheinstetten, Germany

Received: 28 June 2006 / Revised: 22 August 2006 / Accepted: 2 September 2006
Online publication: 5 October 2006

Abstract

The improvements in separation and analysis of complex mixtures by LC-NMR during the last decade have shifted its emphasis from data acquisition to data analysis. For correct data analysis, not only high quality datasets are necessary, but adequate software and adequate databases for semi (or fully)-automated assignments of complex molecules are needed. Only by using NMR, when necessary in combination with MS, the identification of molecules, as present for example in natural products, can be achieved. Here we report on the ongoing efforts required for the construction of an NMR database of flavonoids, implemented for automated assignments of flavonoids. The procedure is demonstrated for a series of flavonoids.

Keywords

Column liquid chromatography
Nuclear magnetic resonance
Mass spectrometry
Automation
Flavonoid database

Introduction

Flavonoids are an important class of secondary metabolites naturally occurring in plants. Flavonoids are flavone-based compounds, and in agreement with Harborne [1], can be classified into 12 subclasses, according to the oxidation level of the central pyran ring (C): anthocyanins, chalcones, aurones, flavones, flavonols, flavanones, dihydrochalcones, proanthocyanidins, catechins, flavan-3,4-diols, biflavonoids and isoflavonoids (Fig. 1). Phenolic compounds occur in plants primarily in a conjugated form. Common substitution patterns of flavonoids include hydroxylation, methoxylation, methylation and/or glycosylation [1].

The interest on studying polyphenols, in particular flavonoids, lies not only in their biological role in plants, but perhaps even more in their potential health benefits for humans. A large number of observational epidemiological studies have shown that specific flavonoid-containing diets can be associated with reduced risks of specific forms of cancer or cardiovascular diseases [2]. Identification of the actual compound in a specific diet responsible for the claimed health effect remains an important bottleneck. The detection, isolation and characterization of low-abundant compounds from complex mixtures depend on an efficient analytical procedure to assure correct identification.

LC-MS is a fast and useful method for the profiling of metabolites present in mixtures. In combination with photo diode array (PDA) and MS/MS information, LC-MS can provide valuable information for identification purposes. By using this combined technology, identification of compounds can be accomplished by testing commercially available standards and by making use of prior biochemical knowledge about the chemical composition of the mixture, either from literature or from dedicated databases [3]. In many situations, especially for complex molecules, LC-PDA-MS/MS may not provide sufficient selectivity for the structure elucidation analysis. In fact, the distinction between possible (dia-)stereoisomers or some constitutional isomers proves to be inconclusive solely by using this LC-PDA-MS/MS technology. A common difficulty is the elucidation of conjugated forms of flavonoids, in which the position of the substitution and/or the nature of the substituent cannot be defined based on LC-PDA-MS/MS data, meaning that

Presented at: Conference of the Hyphenation of Liquid Chromatography-Nuclear Magnetic Resonance Spectroscopy, Liquid Chromatography-Mass Spectrometry and Related Topics, Tuebingen, Germany, March 25–29, 2006.

DOI: 10.1365/s10337-006-0077-6
0009-5893/06/11
© 2006 Friedr. Vieweg & Sohn/GWV Fachverlage GmbH

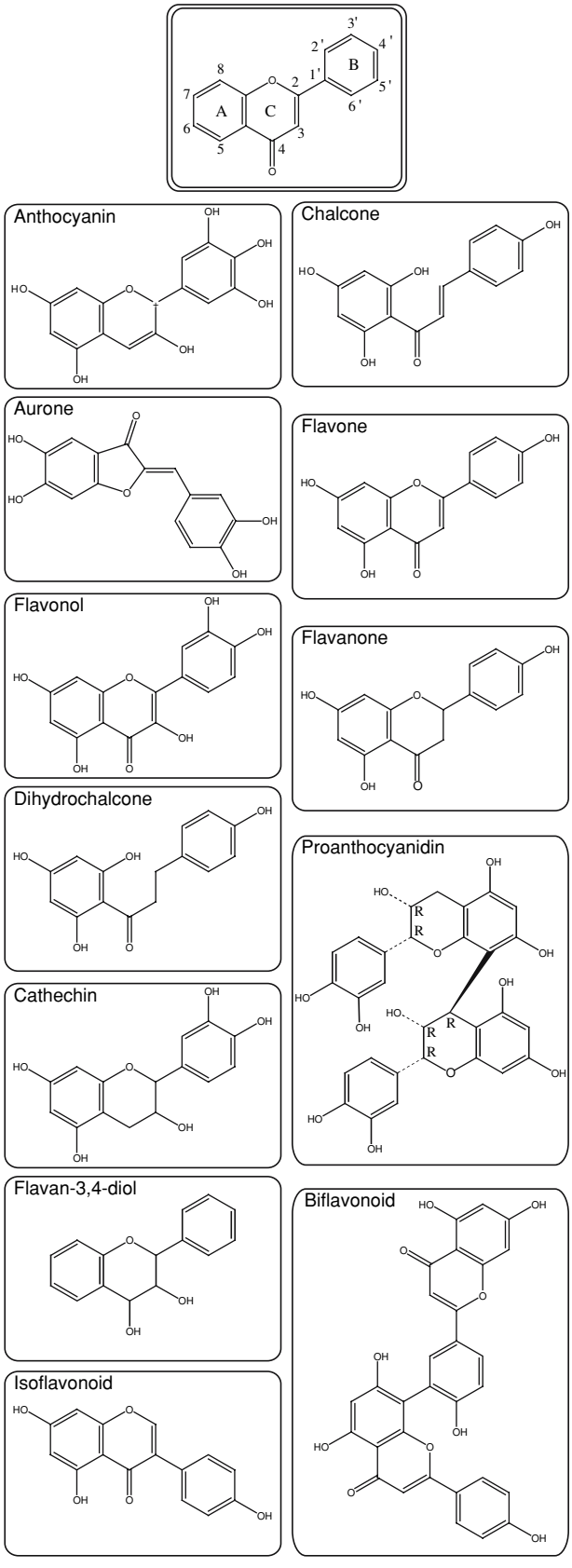


Fig. 1. Structures of examples of molecules in the different flavonoid classes: delphinidin (anthocyanin), naringenin chalcone (chalcone), aureusidin (aurone), apigenin (flavone), quercetin (flavonol), naringenin (flavanone), phloretin (dihydrochalcone), proanthocyanidin B2 (proanthocyanidin), catechin (catechin), flavan-3,4-diol (flavan-3,4-diol), amentoflavone (biflavonoid), genistein (isoflavonoid)

additional NMR data are required. The combination of LC with MS, PDA and NMR is one of the most powerful methods to separate and structurally elucidate unknown compounds from biochemical mixtures [4]. It combines the separation over a wide range of polarities (LC) with photo-spectrometric information (PDA), molecular mass value (MS) and full structural elucidation capabilities (NMR). Specifically, LC-solid phase extraction-NMR (LC-SPE-NMR) [5, 6] and capillary LC-NMR (capLC-NMR) [7] methods, which have been developed recently, improve the isolation and efficiency of identification of compounds present in mixtures.

Given the diversity of secondary metabolites, especially concerning the variety of possible conjugations, the assignment of structures by NMR, and the often tedious analytical isolation, it is essential that reliable NMR-based metabolite databases are constructed. These can facilitate the identification procedure by preventing the isolation of large amounts of material for a full structure elucidation by ^{1}H–^{13}C NMR data. In order to improve the identification of flavonoids present in complex mixtures, we are assembling a database of a wide variety of flavonoids. This database is mainly focused on NMR data of more than 220 flavonoids, acquired under controlled experimental conditions. For all flavonoids, 1D and 2D NMR datasets have been obtained and the ^{1}H and ^{13}C resonances have been assigned. In addition the 1D ^{1}H NMR resonances have been fitted with the PERCH software (http://www.perchsolutions.com, [8]) in order to obtain precise coupling constants. In this study, the precise chemical shifts values and coupling constants as obtained by using PERCH on ^{1}H NMR of 12 related flavonoids: quercetin (**1**) and the derivatives quercetin-4′-*O*-glucoside (**2**) and quercetin-3-*O*-glucoside (**3**), naringenin (**4**) and the derivative naringenin-7-*O*-glucoside (**5**), kaempferol (**6**) and the derivatives kaempferol-3-*O*-rutinoside (**7**) and kaempferol-7-*O*-neohesperoside (**8**), apigenin (**9**), syringetin (**10**) and the derivatives syringetin-3-*O*-galactoside (**11**) and syringetin-3-*O*-glucoside (**12**) are shown. From the data obtained, we evaluate the possibility to identify many flavonoids and derivatives thereof solely on ^{1}H NMR chemical

shifts and ^{1}H-^{1}H NMR coupling constant values, without the need of going into extensive ^{13}C NMR data acquisition routines.

Experimental

Materials and Reagents

The flavonoids **2**, **3**, **5–7**, **8**, **10–12** were purchased from Extrasynthese (France), **1** from Sigma (Germany), **4** from Aldrich (Germany) and **9** from Fluka (Germany). The methanol-d_4 was obtained from Deutero GmbH (Germany) and tetramethylsilane (TMS) from Merck (Germany).

Sample Preparation

A total of 3–5.5 mg of flavonoid was dissolved in 650 µL methanol-d_4. The solutions were homogenized by vortex and for some samples the ultrasonic bath was used to obtain solubility. From the homogenized solution, 600 µL were taken into the NMR tube.

NMR Measurements

NMR measurements were acquired at 300 K using a Bruker Avance 600 spectrometer, proton frequency 600.23 MHz, equipped with a 5 mm TXI probe. Data acquisition was controlled under ICON-NMR version 3.5.6, Bruker XWIN-NMR version 3.5 and Bruker TopSpin version 1.3 (Germany). The 1D ^{1}H spectra were acquired with 65 K data points over a spectral width of 20.028 ppm. The following 2D experiments were recorded: COSY and TOCSY (spectral width 16.0194 ppm in both dimensions; 400 experiments in t_1), J-resolved (spectral width 16.6602 ppm in t_2 and 0.1302 in t_1; 128 experiments in t_1), HSQC (spectral width 16.0194 ppm in t_2 and 185.0601 in t_1; 400 experiments in t_1) and HMBC (spectral width 16.0194 ppm in t_2 and 222.3160 in t_1; 400 experiments in t_1).

Data Analysis

Chemical shifts were referenced to TMS signal (δ = 0 ppm). Bruker TopSpin version 1.3 was used for data conversion and data analysis. PERCH 2005 (Finland) [8] was used for ^{1}H NMR spectral and line shape analysis. AMIX 3.6.8 (Germany) was used for data-integration.

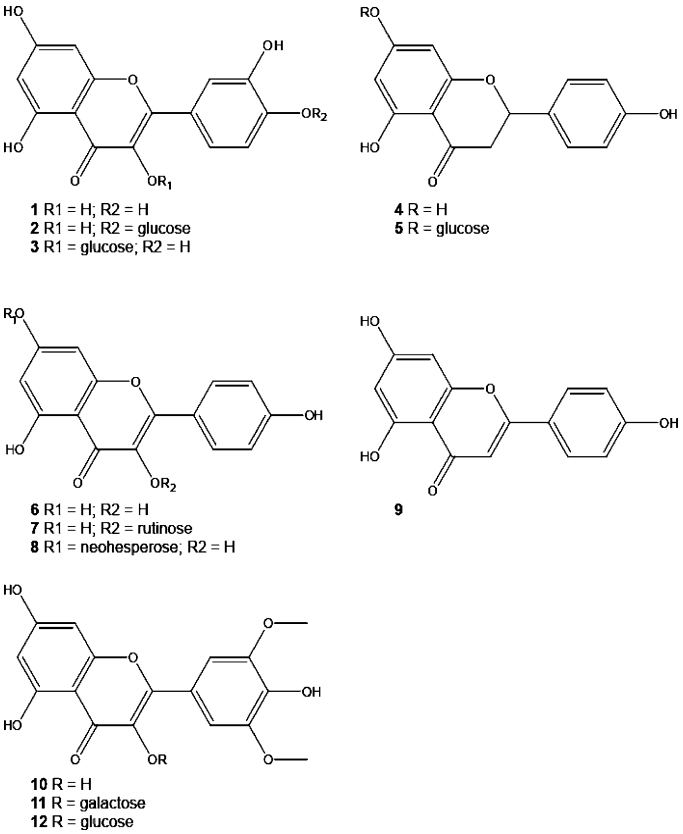


Fig. 2. Structure of flavonoids: **1**, quercetin; **2**, quercetin-4′-*O*-glucoside; **3**, quercetin-3-*O*-glucoside; **4**, naringenin; **5**, naringenin-7-*O*-glucoside; **6**, kaempferol; **7**, kaempferol-3-*O*-rutinoside; **8**, kaempferol-7-*O*-neohesperoside; **9**, apigenin; **10**, syringetin; **11**, syringetin-3-*O*-galactoside; **12**, syringetin-3-*O*-glucoside

Results and Discussion

Measurement and Assignment of the ^{1}H NMR Spectra of the Flavonoids

Assignments of protons and carbons within each measured flavonoid (**1**–**12**) were made by data analysis of the 1D and 2D NMR spectra.

Strategy for the Identification of the Flavonoid and its Sites of Substitution

Glycosylated Flavonoids

As a first example of the subtle but consistent changes on the NMR characteristics of a molecule which occur on substitution, quercetin (**1**) and two conformational isomers **2** and **3** (Fig. 2) have been studied. These two conformational isomers cannot be easily distinguished by MS(-MS), as they are constitutional isomers in which a glucose moiety is placed at two different positions in the flavonoid molecule, 4′ and 3, respectively. However, the ^{1}H-NMR analyses can reveal the identity of these two closely related molecules based on chemical shift data. Table 1 shows the ^{1}H NMR chemical shift values of these two glycosylated quercetin derivatives (**2** and **3**), as well as the resonances of the aglycone quercetin (**1**) itself. The H6 and H8 protons of both **2** and **3** do not shift relative to **1**. In contrast, the H2′, H5′ and H6′ protons of **2** shift in comparison to **1** by 0.03, 0.42 and 0.08 ppm, respectively. The protons of the A- and B-rings of **3** do not shift relative to **1**, indicating that the glucose-moiety is not attached to either one of these two aromatic rings. The large downfield shift of H5′ in **2** relative to **1** indicates a substitution at a *ortho* position to the H5′ proton, i.e., at C4′. The observed shift effects for **2**, especially protons located *ortho* to the substituting group, is consistent throughout other

Table 1. ^{1}H NMR chemical shifts of a selected group of flavonoids (see Fig. 2)

Position	1	2	3	4	5	6	7	8	9	10	11	12
H2				5.33	5.38							
H3a				2.69	2.75				6.59			
H3b				3.11	3.17							
H6	6.18	6.19	6.20	5.88	6.19	6.18	6.21	6.42	6.21	6.19	6.21	6.21
H8	6.39	6.39	6.39	5.89	6.21	6.39	6.41	6.74	6.46	6.42	6.42	6.42
H2′	7.73	7.76	7.71	7.31	7.32	8.08	8.06	8.12	7.85	7.57	7.52	7.57
H3′				6.81	6.81	6.90	6.89	6.91	6.93			
H5′	6.88	7.30	6.87	6.81	6.81	6.90	6.89	6.91	6.93			
H6′	7.63	7.71	7.58	7.31	7.32	8.08	8.06	8.12	7.85	7.57	7.52	7.57
H1″		4.91	5.25		4.96		5.13	5.18			5.47	5.40
H2″		3.55	3.48		3.45		3.42	3.69			3.46	3.82
H3″		3.51	3.42		3.43		3.33	3.63			3.45	3.57
H4″		3.43	3.34		3.38		3.25	3.40			3.30	3.84
H5″		3.48	3.22		3.45		3.41	3.54			3.25	3.48
H6a″		3.93	3.57		3.87		3.80	3.70			3.75	3.60
H6b″		3.74	3.71		3.68		3.38	3.92			3.57	3.66
H3′5′M										3.93	3.93	3.94

The ^{1}H NMR chemical shifts values of the rhamnoside moiety of compounds **7** and **8** are not shown
M methoxy

analogous examples in our enlarged database. The absence of substantial chemical shift effects for the aromatic ring protons of **3** relative to **1** indicates that the electronic configuration of the backbone structure does not change. This suggests that the glucose moiety should therefore be attached to the 3-OH group in the C-ring.

From our database, a small selection of related glycosylated forms of the aglycones **4**, **6** and **10** are shown in Fig. 2 (**5**, **7**, **8**, **11**, **12**). The ^{1}H NMR chemical shift values of the glycosylated forms of these flavonoids and their aglycones (Table 1) indicate that the protons located *ortho* to the *O*-glycosylation position shift to higher ppm values by at least 0.24 ppm. The chemical shift values of the H6 and H8 protons of structurally related molecules (**1**, **2**, **3**, **6**, **7**, **9**, **10**, **11**, **12**), in the absence of substitution on C5 or on C7, are virtually identical. The presence of a saturated bond between C2 and C3, as present in naringenin (**4**) and derivatives thereof (for example, **5**), does shift the H6 and H8 protons to lower ppm values by 0.3 ppm. This observation indicates a change in backbone structure, as well as in electronic configuration.

Methoxylated Flavonoids

In addition to the systematic changes of the shifts observed of glycosylated flavonoids, also systematic changes are, for example, observed on *O*-methylation. Our observations are consistent with the recent publications of Lambert et al. [9] in which isoflavonoids from *Smirnowia iranica* have been identified by LC-SPE-NMR and of Kim et al. [10], in which a complete set of ^{1}H NMR assignments of flavonol derivatives have been obtained. *O*-methylation of phenol-rings of flavonols [10] or *O*-methylation of *ortho*-disubstituted phenols [9] results in higher ppm values for protons located at the *ortho*, *meta* or *para* position. The *O*-methylation-induced-chemical shifts can therefore be used for the identification of *O*-methylated flavonoids.

Strategy for Identification of the Type of Glycoside in a Glycosylated Flavonoid

The distinction between different conformations of equal-mass-sugar moieties cannot be readily achieved by electrospray (ESI)-MS. This implies that glycosylated flavonoids with either one or two of the most occurring monosaccharides in plants, such as glucose and galactose, cannot be discriminated by LC-MS-MS, as these two moieties have identical mass. Discrimination between a glucosylated and a galactosylated flavonoid should therefore be done by NMR-based identification. We expect that identification is feasible exclusively based on 1D ^{1}H NMR data, providing that good quality spectra are obtained. However, in contrast to the possibilities for identification of the flavonoid backbone based on the ^{1}H chemical shifts, the identification of the carbohydrate moiety cannot be easily performed by analysis of chemical shift values, as described above. But from careful analyses of more than 25 glucosylated flavonoid derivatives (data not shown), it can be seen that the coupling constant values in the glucopyranoside-ring are very similar for identical through-bond couplings. The $^3J_{H-H}$ coupling constants obtained from evaluating the NMR data of the glucosylated flavonoids **2**, **3**, **5**, **12** using PERCH were $^3J_{H1''-H2''}$ (7.7 ± 0.3) Hz, $^3J_{H2''-H3''}$ (9.3 ± 0.2) Hz, $^3J_{H3''-H4''}$ (9.0 ± 0.1) Hz and $^3J_{H4''-H5''}$ (9.8 ± 0.1) Hz (Table 2). It is well known that these coupling constants are dependent on the dihedral angle and the substitution pattern in the sugar ring [11]. Almost no change in the values of the coupling constants of the different glucopyranosides studied was observed, implying that all these sugar moieties are in the same (minimized) energy conformation, in the solvent used (methanol-d_4). This is an encouraging observation, as it indicates that the type of sugar moiety present in a flavonoid molecule can be obtained from detailed analysis of its ^{1}H NMR splitting patterns. In our opinion, the software tool PERCH [8] best suits the analyses of complex ^{1}H NMR splitting patterns, even in the presence of strong coupling and overlapping signals.

As an example of two closely related glycosylated molecules with the same molecular formula (and hence these molecules can not be discriminated by MS, as both have the same molecular mass, 508 Da), the ^{1}H-NMR regions of the two sugar moieties of **11** and **12** are shown (Fig. 3). The complexity of the splitting patterns is due to the partially

Table 2. ^{1}H NMR coupling constants of a selected group of flavonoids (see Fig. 2)

Coupling	1	2	3	4	5	6	7	8	9	10	11	12
^{4}J(H6, H8)	2.1	2.1	2.1	2.2	2.3	2.1	2.1	2.1	2.1	2.0	2.1	2.1
^{3}J(H2′, H3′)				8.4	8.4	8.8	8.7	8.8	8.7			
^{3}J(H5′, H6′)	8.5	8.7	8.5	8.4	8.4	8.8	8.7	8.8	8.7			
^{4}J(H2′, H6′)	2.2	2.2	2.2	2.6	2.5	2.6	2.4	2.5	2.5	2.0	2.0	2.0
^{4}J(H3′, H5′)				2.4	2.4	2.7	2.5	2.5	2.4			
^{3}J(H1″, H2″)		7.8	7.9		8.0		7.4	7.8			7.8	7.7
^{3}J(H2″, H3″)		9.4	9.3		9.5		9.2	9.2			9.6	9.3
^{3}J(H3″, H4″)		9.0	9.0		8.9		9.0	9.2			3.4	8.9
^{3}J(H4″, H5″)		9.8	9.8		9.8		9.9	9.8			1.1	9.8
^{3}J(H5″, H6a″)		2.3	2.4		2.0		1.8	2.3			6.5	2.3
^{3}J(H5″, H6b″)		5.7	5.4		5.0		6.1	5.9			5.6	5.6
^{2}J(H6a″, H6b″)		−12.1	−11.9		−12.1		−11.2	−12.3			−11.3	−11.9
^{2}J(H3b, H3a)				−17.1	−17.2							
^{3}J(H3b, H2)				13.0	12.9							
^{3}J(H3a, H2)				3.0	3.1							

The ^{1}H NMR coupling constant values of the rhamnoside moiety of compounds **7** and **8** are not shown

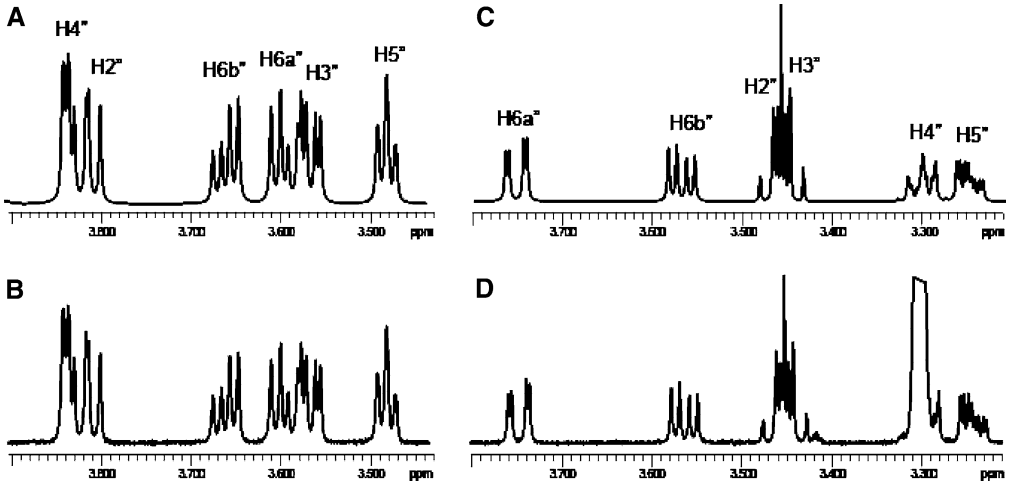


Fig. 3. The spectra of syringetin-3-*O*-galactoside (**a**) and syringetin-3-*O*-glucoside (**c**) using PERCH with preset coupling constants for the different sugar pyranoside rings and adjusting the ^{1}H resonances of H2″, H3″, H4″, H5″, H6a″ and H6b″ to the most optimal chemical shift position. The sugar region of the measured ^{1}H NMR spectra of syringetin-3-*O*-galactoside (**b**) and syringetin-3-*O*-glucoside (**d**) are depicted. Note the solvent resonance of methanol (3.30 ppm), still visible in **d**, which in **c** is absent using PERCH

overlapping resonances and also to the presence of strong coupling effects. Nevertheless, when fitting the NMR spectra using the PERCH software programme, all coupling constants are adequately obtained (Fig. 3) and the difference between a glucopyranose and a galactopyranose can be readily observed. In our fitting procedure, a peak-top interpolation is firstly performed, including a line-shape analysis, followed by deconvolution with total-line-shape fitting. When fixing the coupling constants to specific preset values, as discussed above, it is possible, in an iterative manner, to obtain a convergent fit of the observed ^{1}H NMR spectrum, resulting in the ^{1}H NMR chemical shift values for the glucopyranose or galactopyranose moieties. When exchanging H2″ to H3″ in Fig. 3d (**12**), for example, the PERCH fit does not adequately converge, indicating that the proton resonances have not been correctly assigned.

Applicability of Fitting Strategies for Identification Purposes

Identification of conjugations present in the flavonoid backbone using the procedure described requires some form of purification, either by using LC-NMR or preferably by using LC-SPE-NMR [4]. Peaks separated in a LC column can contain more than one compound, but as first shown by Exarchou et al. [5], even with two or three compounds trapped into the same SPE cartridge, the NMR spectra can be of the required quality. Subsequent fitting of the data, as proposed in this study, can result in the correct identification, thereby avoiding time consuming ^{1}H–^{13}C HSQC or ^{1}H–^{13}C HMBC data acquisition and analysis. We expect that when using LC-SPE-NMR with cryoprobes with flow insert, the identification of flavonoids can be achieved in the 500–800 ng region in a 30 min-^{1}H NMR-run. This amount of material is much less than when one has to rely on identification based on 2D ^{1}H–^{13}C HSQC and ^{1}H–^{13}C HMBC data [5].

Building up a NMR Database of Flavonoids

The incorporation of the flavonoid ^{1}H NMR chemical shifts into a searchable database will provide predictive ^{1}H NMR

values for conjugated flavonoids facilitating identification. Moreover, the small but systematic changes observed can be of value for theoretical chemists who are developing programs for prediction of ^{1}H NMR chemical shift values [12]. Over the last decade, these theoretical prediction programs have steadily improved in performance but, as yet, are not able to predict the ^{1}H NMR chemical shift values for flavonoids or related complicated biological molecules with the precision needed for identification, solely based on ^{1}H NMR chemical shift values. All our data are being incorporated into AMIX (MOL files, ^{1}H and ^{13}C chemical shifts) and will be published in more detail in the near future. Its usage in complex mixture analysis will be demonstrated.

Conclusions

The ^{1}H NMR chemical shift values of flavonoids show systematic differences depending on the substitution pattern. These systematic differences apply to the protons of the backbone moiety. The resonances of the sugar moieties in glycosylated flavonoids are too scattered to be of use in direct identification. However, as we have observed, the coupling constants are extremely predictable and through fitting the ^{1}H resonances with preset coupling constants in PERCH, discrimination between the type of sugar, for example galactopyranoside or glucopyranoside, can be readily obtained. As the ^{1}H NMR spectrum of glycosylated flavonoids is crowded with resonances in the 2.8–4.0 ppm region (with often strong coupling effects), it is essential to isolate the flavonoid under study to a reasonable degree of homogeneity, preferably through LC-SPE-NMR.

References

1. Harborne JB (1980) In: Bell EA, Charlwood BV (eds) Encyclopedia of Plant Physiology, Secondary Plant Products, vol 8. Springer, Berlin Heidelberg New York
2. Ross JA (2002) Annu Rev Nutr 22:19–34
3. Moco S, Bino RJ, Vorst O, Verhoeven HA, de Groot J, van Beek TA, Vervoort J, Ric de Vos CH (2006) Plant Physiol 141:1205–1218
4. Wolfender J-L, Ndjoko K, Hostettmann K (2003) J Chromatogr A 19:437–455
5. Exarchou V, Godejohann M, van Beek TA, Gerothanassis IP, Vervoort J (2003) Anal Chem 75:6288–6294
6. Exarchou V, Krucker M, van Beek TA, Vervoort J, Gerothanassis IP, Albert K (2005) Magn Reson Chem 43:681–687
7. Krucker M, Lienau A, Putzbach K, Grynbaum MD, Schuler P, Albert K (2004) Anal Chem 76:2623–2628
8. Laatikainen R, Niemitz M, Weber U, Sundelin J, Hassinen T, Vepsäläinen J (1996) J Magn Reson Ser A 120:1–10
9. Lambert M, Staerk D, Hansen SH, Sairafianpour M, Jaroszewski JW (2006) J Nat Prod 68:1500–1509
10. Kim H, Moon B-H, Ahn J-H, Lim Y (2006) Magn Reson Chem 44:188–190
11. Haasnoot CAG, De Leeuw FAAM, Altona C (1980) Tetrahedron 36:2783–2792
12. Abraham RJ, Byrne JJ, Griffiths L, Koniotou R (2005) Magn Reson Chem 43:611–624