ELSEVIER

Carbohydrate Research 302 (1997) 67–78

CARBOHYDRATE RESEARCH

Isolation and structure elucidation of a highly haemolytic saponin from the Merck saponin extract using high-field gradient-enhanced NMR techniques

Corinne Delay [b], José A. Gavin [c], André Aumelas [a], Pierre-Antoine Bonnet [b], Christian Roumestand [a,*]

[a] *Centre de Biochimie Structurale, CNRS-UMR 9955 et INSERM-U414, Université de Montpellier I, Faculté de Pharmacie, 15 Avenue Charles Flahault, F-34060 Montpellier, France*

[b] *Laboratoire de Chimie Organique, Université de Montpellier I, Faculté de Pharmacie, 15 Avenue Charles Flahault, F-34060 Montpellier, France*

[c] *Instituto Universitario Bio-Organica "Antonio Gonzalez", Carretera La Esperanza 2, E-38206 La Laguna (Tenerife), Spain*

Received 26 December 1996; accepted 19 March 1997

Abstract

Saponins SAPO50 and SAPO30, of which SAPO50 is highly haemolytic, have been isolated from the commercial Merck Saponin. Their structures have been determined exclusively by high-field gradient-enhanced NMR methods. The ^{1}H and ^{13}C NMR spectra of these saponins in pyridine–deuterium oxide have been assigned by homonuclear and heteronuclear correlation experiments. Anomeric configurations were obtained by combined use of $^{1}J_{CH}$, $^{3}J_{H\text{-}1,H\text{-}2}$, and 1D-NOESY data. Sugar residues were identified by use of $^{3}J_{HH}$ values obtained from their subspectra recorded using an optimized 1D-*z*-TOCSY sequence. Linkage assignments were made using the ge-HMBC and 1D-NOESY spectra. This study shows that SAPO50 represents a hitherto undescribed saponin with the following structure: 3-*O*-β-D-xylopyranosyl-(1 → 3)-[β-D-galactopyranosyl-(1 → 2)]-β-D-glucuronopyranosyl gypsogenin 28-*O*-(6-deoxy-β-D-glucopyranosyl)-(1 → 4)-[β-D-xylopyranosyl-(1 → 3)-β-D-xylopyranosyl-(1 → 4)]-α-L-rhamnopyranosyl-(1 → 2)-β-D-fucopyranoside. SAPO30, however, corresponds to a saponin previously described [D. Frechet, B. Christ, B. Monegier du Sorbier, H. Fischer, and M. Vuilhorgne, *Phytochemistry*, 30 (1991) 927–931]. © 1997 Elsevier Science Ltd.

Keywords: Haemolytic saponins; 2D-NMR spectroscopy; Gradient-enhanced NMR experiments; Multiplicity-edited ge-HSQC

* Corresponding author.

PII S0008-6215(97)00101-8

1. Introduction

Saponins are steroid or triterpene glycosides that have the distinctive property of forming a soapy lather in water. This group of natural substances possess a broad spectrum of biological properties [1]; in particular, they have been described as potent haemolytic agents [2,3]. The commercial Merck Saponin (E. Merck No. 7695), also known as Saponin Pure White, is a crude saponin fraction obtained from roots and rhizomes of *Gypsophyla paniculata*. Due to its high haemolytic properties, it has been extensively used in the past as a standard for haemolytic tests in most saponin determinations. In order to identify the saponins responsible for the high-haemolytic activity, we have used chromatographic methods to separate this crude product. A highly haemolytic fraction was analyzed further and two major compounds were isolated by reverse-phase HPLC: SAPO50 and SAPO30, which possess 25% and 1%, respectively, of the total haemolytic activity of the crude extract. The structure of such molecules composed of one or several sugar chains linked to an aglycone is usually determined by derivatization followed by degradation [2,4] or by ^{13}C and/or ^{1}H spectroscopy [5,6]. Surprisingly, there are few examples in the literature of structure elucidation of complex saponins using only homo- (e.g. Ref. [7]) and/or hetero-nuclear NMR spectroscopy (e.g. Refs [8–10]), without recourse to chemical degradation. Even though NMR spectroscopy of very large biomolecules (proteins, nucleic acids) in solution is now commonly used for assignment purposes and structural determinations, oligosaccharides of natural or synthetic origin, even of moderate complexity, often present proton spectra (and sometimes carbon spectra) not amenable to direct analysis. Most of the non-anomeric protons appear in the 3–5 ppm region, leading to excessively overcrowded spectra. In this paper, we describe the sole use of multidimensional ^{1}H and ^{13}C (natural abundance) NMR spectroscopy [mainly gradient-enhanced (ge-) techniques] at high field (600 MHz for proton) to solve the structure of SAPO30 and SAPO50. We found that SAPO30 corresponds to a saponin previously described by Frechet et al. as 3-*O*-β-D-xylopyranosyl-(1 → 3)-[β-D-galactopyranosyl-(1 → 2)]-β-D-glucuronopyranosyl gypsogenin 28-*O*-β-D-glucopyranosyl-(1 → 3)-[β-D-xylopyranosyl-(1 → 4)]-α-L-rhamnopyranosyl-(1 → 2)-β-D-fucopyranoside [8]. Since our data are in complete agreement with the reported data, we restrict this report to the structure elucidation of SAPO50.

2. Results and discussion

Identification of the aglycone moiety.—Since the proton and carbon resonances of the aglycone moiety are well resolved at high field (600 MHz for proton, 125 MHz for carbon), their assignment, based on proton-detected heteronuclear experiments, was relatively straightforward (data not shown).

Protonated carbon resonances, as well as the carbon multiplicity, were determined using a gradient-enhanced (ge-) multiplicity-edited PEP-HSQC experiment [1]. Compared to the traditional phase-cycled HSQC experiment, the use of pulse-field gradient to select the desired coherence-transfer pathway has the considerable advantage of effectively discarding undesired magnetization, such as that from the ^{12}C-bonded proton ($^{1}H-^{12}C$), without the need of preparation period (e.g. the BIRD [11] pulse sequence). Moreover, the sensitivity loss associated to conventional gradient-enhanced experiments, due to the selection of a single specific coherence-transfer pathway [12–14], can be easily balanced by the use of the PEP (Preservation of Equivalent Pathways) methodology [15,16] shown to improve the signal-to-noise ratio [17,18]. Furthermore, the implementation of an editing block in the PEP-HSQC sequence, as recently proposed by Parella et al. [19] [2], allows us to elucidate the carbon multiplicity. Thus, of the twenty-two signals in the aglycone spectrum, corresponding to protonated carbons, ten were identified as secondary carbon atoms based on their negative sign. Six of the twelve remaining positive signals were unequivocally assigned to methyl groups, and an aldehydic group was identified based on its low-field shift (208.7 ppm).

A 2D-HSQC-TOCSY [20] experiment recorded with a mixing time of 80 ms gives the connectivities between the different protonated carbons. In this case, the conventional phase-cycled HSQC [21] pulse sequence was used, with the addition of pulse-field gradients as spoils in order to destroy the $^{1}H-^{12}C$ magnetization, thus avoiding the use of the BIRD sequence. The simple implementation of these purges on HSQC-type experiments [17] was the reason for their preference over the HMQC-type [22] homologue pulse sequences. In combination with the results obtained from a high-resolution DQF-COSY

[1] T. Parella, F. Sanchez-Ferrando, and A. Virgili, personal communication, in preparation.

[2] T. Parella, F. Sanchez-Ferrando and A. Virgili, personal communication, in preparation.

[23] experiment, which yields the complete proton coupling network, all the proton and the protonated carbon resonances could be unambiguously assigned (data not shown).

The quaternary carbons were identified using the assigned protonated carbon resonances by means of a long-range proton–carbon correlation experiment (data not shown). This latter experiment was recorded using the gradient-enhanced version of the heteronuclear multiple-bond correlation (HMBC) experiment [24–26]. Here again, the selection of coherence pathways by an adequate combination of B_0 pulse-field gradients effectively discards all subtraction artefacts usually observed in the phase-cycled version of this experiment.

By comparison with literature data [27], the aglycone was identified as gypsogenin. The ^{1}H and ^{13}C assignments are listed in Table 1.

Identification and NMR assignments of sugar residues.—The identification of the sugar residues and their connectivities proved more challenging, given the extensive overlap of their NMR signals, even at very high field. The identification of the sugar residues, based on their anomeric protons, represented the first step. Proton subspectra of the various carbohydrate moieties were obtained from the rows corresponding to their anomeric proton resonances or methyl groups in a *z*-TOCSY [28–30] experiment with a mixing time of 150 ms (Fig. 1). In this manner, eight sugar units [individual sugars are indicated by bold capital letters (**A**–**H**), following their chemical shift in the 1D proton spectrum] were identified in SAPO50 (seven in SAPO30) based on their characteristic proton signals in the 3.5–5 ppm region. Note that the additional proton resonance in the anomeric region previously has been assigned to the vinylic proton (H-12) of the gypsogenin. This result was strongly supported by the presence of eight ^{13}C resonances in the HSQC spectrum in the 90–105 ppm range (C-12: 119.1 ppm). Correlations between three methyl resonances and sugar resonances indicate the presence of three methylated sugar units: two were unambiguously assigned to sugars **A** and **G**, by comparison of the subspectra extracted either from the corresponding anomeric protons or from the methyl group. The third could not be determined at this step, as only one correlation to this CH_3 group was observed in the TOCSY spectrum, which cannot lead to an unequivocal assignment.

The problem, generally faced when studying saccharides, is the determination of the direct connectivities between the proton resonances belonging to the same sugar unit. Due to the severe overlap in the 3.5–5 ppm region of the proton spectrum, almost only H-1,H-2 connectivities (from anomeric protons), and ^{3}H-6,H-5 connectivities (from methyl groups, ^{3}H-6 stands for the methyl groups at position 6) can be unambiguously identified from a DQF-COSY experiment. To address this problem, stepwise RELAY experiments have been proposed [31], but this approach implies the a priori knowledge of the ^{3}J coupling constants for the different protons of a same sugar unit in order to optimize the relay-transfer times. Moreover, this approach fails in the case of strongly coupled systems, as observed in residue **H**. The strategy based on heteronuclear experiments, namely the multiplicity-edited ge-PEP-HSQC spectrum and 2D-HSQC-TOCSY spectra, circumvents the problems described (Fig. 2). The HSQC spectrum

Table 1
^{13}C and ^{1}H NMR data (δ in ppm) [a] for the aglycone moiety of SAPO50 (and SAPO30) in 1:1 pyridine-d_5–D_2O. The ^{13}C chemical shifts of boundary carbons with the sugar chains are italicized.

Position	^{13}C	^{1}H
C-1/H-1	35.0	1.60/1.62
C-2/H-2	21.8	2.33/2.02
C-3/H-3	*82.4*	4.17
C-4	52.7	
C-5/H-5	45.1	1.49
C-6/H-6	17.4	1.49/1.14
C-7/H-7	29.4	1.66/1.52
C-8	37.7	
C-9/H-9	44.6	1.73
C-10	32.6	
C-11/H-11	20.5	1.93/1.89
C-12/H-12	119.1	5.44
C-13	140.7	
C-14	39.0	
C-15/H-15	25.4	1.86/1.45
C-16/H-16	20.1	2.11/1.85
C-17	45.3	
C-18/H-18	38.9	3.06
C-19/H-19	43.2	1.75/1.23
C-20	27.0	
C-21/H-21	30.7	1.32/1.21
C-22/H-22	29.2	1.90/1.72
C-23/H-23	208.7	9.8
C-24/H-24	7.6	1.5
C-25/H-25	12.6	0.93
C-26/H-26	14.2	0.99
C-27/H-27	22.8	1.28
C-28	*174.0*	
C-29/H-29	30.0	0.98
C-30/H-30	20.5	0.93

[a] Chemical shifts with the highest field signals of pyridine-d_5 as reference (7.19 ppm for ^{1}H and 123.5 ppm for ^{13}C).

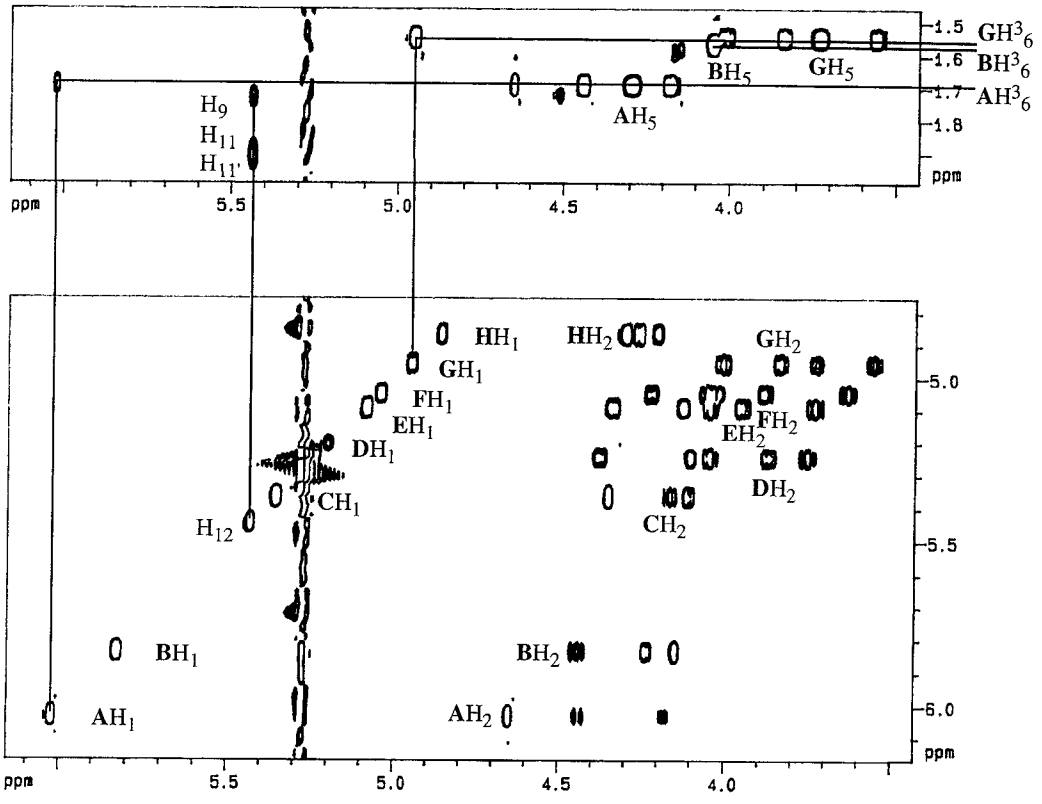


Fig. 1. Slices extracted from the *z*-TOCSY spectrum (mixing time 150 ms) of SAPO50 showing (Top) the correlations obtained from the methyl resonances and (Bottom) the correlations obtained from the anomeric resonances. Sugar units are labelled by bold capital letters (**A–H**, H^3-6 stands for the methyls groups at position 6), following their anomeric proton chemical shift. Vertical lines connect the similar subspectra obtained from both the methyl groups and the anomeric resonances. The resonances of protons H-2 and H-5 reported in this spectrum have been identified by their direct connectivities in a DQF-COSY spectrum.

leads to direct (one-bond) $^1H-^{13}C$ correlations, whereas a HSQC-TOCSY recorded with a short mixing time (16 ms) gives additionally the vicinal correlation. The result of this experiment is equivalent to that of a 2D-HMQC-COSY [22]; however, the observed correlation cross-peaks are in phase preventing the accidental cancelling of overlapping peaks. Moreover, in our hands, this experiment appeared considerably more sensitive than the corresponding HMQC-COSY. Note that with the small mixing time used, the three-bond connectivity generally vanishes when the ^3J proton is small (< 3 Hz), thus giving semi-quantitative information on the relative stereochemistry of the two bound carbons. This correlation is usually present in the HSQC-TOCSY recorded with a longer mixing time (80 ms). In this latter experiment, correlation peaks from the anomeric proton could be traced to all ^{13}C resonances belonging to the same residue (Fig. 2).

Finally, to identify the carbohydrate unequivocally, the relative stereochemistry of each asymmetric centre was determined based on the vicinal coupling constants. Since all anomeric resonances are well separated at this frequency, 1D-*z*-TOCSY experiments (150 ms mixing time) [32] were used. Subramanian and Bax have shown that the use of the *z*-filtering [33] technique allows for the acquisition of pure-phase subspectra [34], which is mandatory when measuring coupling constants. The very robust DANTE-Z [35,36] sequence was used for the frequency selection, allowing an easy implementation of the 1D-*z*-TOCSY experiment. A minor but excessively efficient modification was introduced in our previously published sequence, consisting in a weak pulse-field gradient (or, alternatively, a homospoil pulse) in order to purge spurious coherences at the end of the mixing period (Fig. 3). Due to the high digital resolution of the 1D subspectra obtained, coupling constants could be determined with sufficient accuracy (Fig. 4). Additional information on the configuration has been extracted from 1D-NOESY experiment [32] obtained by selective inversion of the anomeric protons.

The details of the application of this strategy to the

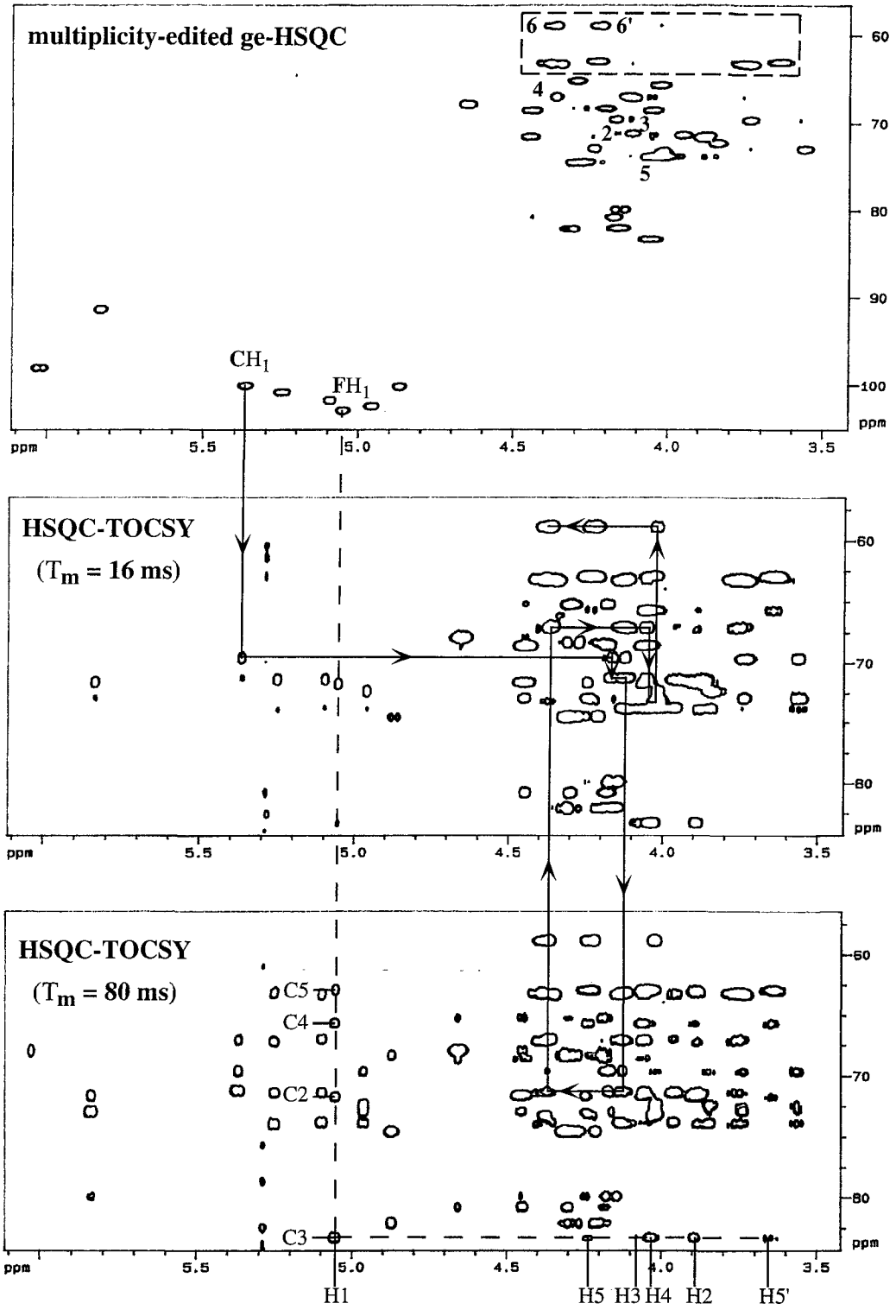


Fig. 2. Sugar region of the multiplicity-edited ge-HSQC and HSQC-TOCSY (mixing times 16 and 80 ms) spectra of SAPO50. The multiplicity-edited ge-HSQC shows the direct $^{13}C-^{1}H$ correlation. Positive and negative peaks are not differentiated in the plot, but the negative signals corresponding to methylenic carbons are indicated in the dotted frame. The HSQC-TOCSY with 16 ms essentially displays direct and vicinal correlations, whereas complete proton or carbon subspectra can be extracted from the HSQC-TOCSY recorded with a 80 ms mixing time, as indicated for residue **F** (dotted line). The walk from anomeric ^{13}C and ^{1}H resonances across the full subspectra of residue **C** is indicated by the dark arrowed line. The corresponding direct connectivities are labelled in the HSQC spectrum.

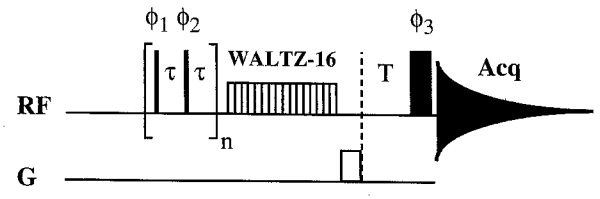


Fig. 3. Pulse sequence of the 1D-z-TOCSY experiment. The DANTE-Z sequence is used for frequency selection. It allows for the inversion of a selected proton multiplet only in odd-numbered scans. Correspondingly, data are added and subtracted by phase alternation of the receptor phase. The isotropic mixing sequence (WALTZ-16) is applied immediately after the DANTE-Z block, followed by a very weak pulse-field gradient (1 ms duration and 0.5 G/cm strength, corresponding to 1% of the maximum amplifier power) or, alternatively, a homospoil pulse which destroys all unwanted transverse magnetization components. The delay time T which follows has been adjusted (6 ms) to minimize perturbation of the last pulse in the sequence by the gradient. Pulses have been phase-cycled as: $\phi_1 = 2x$, $2(-x)$; $\phi_2 = x, -x, -x, x$; $\phi_3 = 4x$, $4(-x)$, $4y$, $4(-y)$; Acq = $2(x, -x)$, $2(-x, x)$, $2(y, -y)$, $2(-y, y)$. Since, the isotropic mixing is applied on z magnetization, the phase of the WALTZ-16 sequence is not critical: we obtained best results when using the same phase cycle as ϕ_1. Note that only the first four steps are needed for the DANTE-Z selection, the additional steps (CYCLOPS) help to reduce quadrature artefacts.

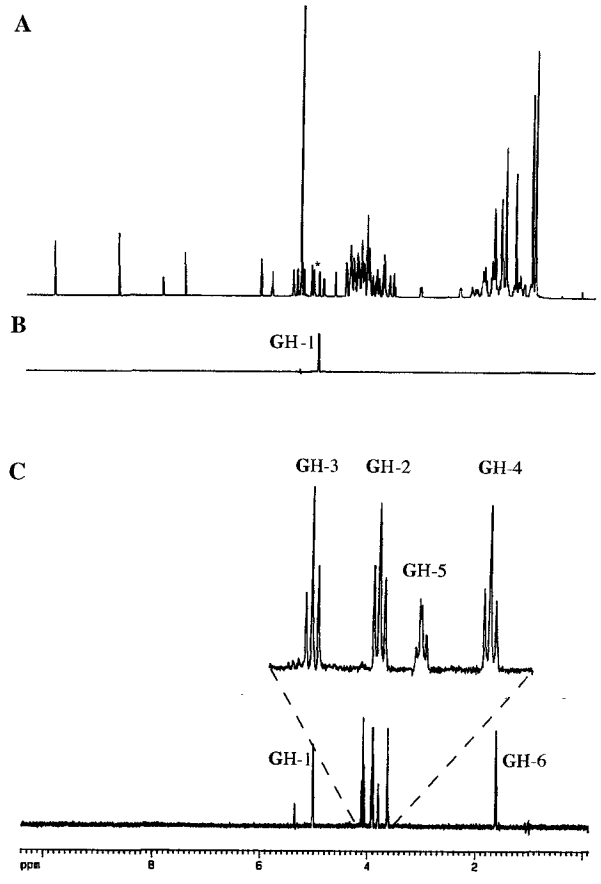


Fig. 4. 600-MHz spectra of SAPO50 in 1:1 pyridine-d_5–D_2O at 320 K. (A) Regular 1D spectrum of SAPO50, recorded with 32 scans (4 dummy scans) and repetition rate of 3 s. The frequency of the anomeric proton of residue **G** is indicated with an asterisk. (B) Selection of the anomeric resonance of residue **G** by the DANTE-Z sequence with the following parameters (37): θ (nutation value of the pulses used in the 180° DANTE train) = 0.3° (corresponding to π pulse length of 1.5 ms), τ (precession delay) = 100 μs, $n = 300$. Four dummy scans and thirty-two scans were acquired, as in (A). (C) Subspectrum of residue **G** obtained with the 1D-z-TOCSY sequence (128 scans, 4 dummy scans) described in Fig. 3, with a mixing period of 150 ms. Inset: a zoom on the sugar-ring proton resonances of this residue shows the excellent phase purity of the subspectrum obtained.

eight monosaccharides constitutive of SAPO50 are discussed below. All NMR data are collected in Table 2 (^{13}C chemical shifts), Table 3 (^{1}H chemical shifts), and Table 4 (^{1}H coupling constants).

Residue A.—The anomeric proton of residue **A** shows only three cross-peaks in the TOCSY spectrum (τ_m 150 ms), but the complete subspectra can be traced from methyl proton AH^3-6 (**A** stands for residue **A**, H^3-6 for the methyl group at position 6) or from the well-resolved AH-2 proton in the same spectrum (not shown). Inspection of the DQF-COSY spectrum shows that the AH-2 and AH-5 resonances could be unequivocally assigned. Since the AH-1, AH-2, and AH^3-6 resonances are well resolved in the 1D spectrum, the vicinal coupling constants $^3J_{H\text{-}1,H\text{-}2}$, $^3J_{H\text{-}2,H\text{-}3}$, and $^3J_{H\text{-}6,H\text{-}5}$ have been accurately measured. $^3J_{H\text{-}3,H\text{-}4}$ and $^3J_{H\text{-}4,H\text{-}5}$ were measured on the 1D-z-TOCSY spectrum obtained from selective excitation of AH-2. Along with $^1J_{CH}$ (171 Hz) and $^3J_{H\text{-}1,H\text{-}2}$ and $^3J_{H\text{-}2,H\text{-}3}$ (1.9 and 4.0 Hz, respectively), residue **A** was assigned as α-rhamnopyranoside since it is known that, in general, $^1J_{CH} < 165$ Hz and $^3J_{H\text{-}1,H\text{-}2} > 5$ Hz indicate a β configuration, and $^1J_{CH} > 165$ Hz and $^3J_{H\text{-}1,H\text{-}2} < 5$ Hz an α configuration [37,38]. The well-resolved assignments of ^{13}C resonances were obtained from HSQC and 2D-HSQC-TOCSY experiments.

Residue B.—In the TOCSY spectrum, the anomeric proton of residue **B** shows three cross-peaks. The inspection of the 2D-HSQC-TOCSY spectra indicates that this residue is a methylated sugar which allows us to identify the non-assigned methyl group in the TOCSY spectrum. In this manner, all proton and carbon resonances of residue **B** were clearly identified. In the TOCSY spectrum, the methyl group

Table 2
^{13}C NMR data (δ in ppm) [a] for the saccharides of SAPO50 in 1:1 pyridine-d_5–D_2O. The ^{13}C chemical shifts of boundary carbons in substituted residues are italicized

Residue	1	2	3	4	5	6
α-Rha*p* (**A**)	98.0 (171) [b]	68.0	68.5	*80.3*	65.2	15.0
β-Fuc*p* (**B**)	91.4 (163) [b]	*71.6*	72.9	*79.9*	68.7	13.8
β-Gal*p* (**C**)	100.0 (160) [b]	69.6	71.2	67.1	73.3	58.9
β-Xyl *p* (**D**)	100.7 (163) [b]	71.3	73.8	67.1	63.3	
β-Xyl *p* (**E**)	101.7 (161) [b]	71.3	73.8	67.1	63.3	
β-Xyl *p* (**F**)	102.8 (158) [b]	71.6	*83.3*	65.6	63.0	
β-6-Deoxy-Glc*p* (**G**)	102.4 (157) [b]	72.3	73.8	72.9	69.6	14.9
β-GlcA *p* (**H**)	100.1 (158) [b]	*74.6*	*82.0*	68.3	74.5	172.4

[a] Chemical shifts with the highest field signals of pyridine-d_5 as reference (123.5 ppm).
[b] $^1J_{CH}$ in Hz.

Table 3
^1H NMR data (δ in ppm) [a] for the saccharides of SAPO50 in 1:1 pyridine-d_5–D_2O

Residue	1	2	3	4	5eq	5ax	6
α-Rha*p* (**A**)	6.01	4.65	4.44	4.17		4.29	1.71
β-Fuc*p* (**B**)	5.82	4.44	4.23	4.15		4.04	1.57
β-Gal*p* (**C**)	5.36	4.16	4.12	4.36		4.01	4.37/4.23
β-Xyl *p* (**D**)	5.24	3.87	4.04	4.11	4.38	3.76	
β-Xyl *p* (**E**)	5.08	3.96	4.04	4.13	4.34	3.73	
β-Xyl *p* (**F**)	5.04	3.88	4.06	4.02	4.22	3.63	
β-6-Deoxy-Glc*p* (**G**)	4.95	3.84	4.01	3.55		3.73	1.55
β-GlcA *p* (**H**)	4.87	4.30	4.31	4.21		4.29	

[a] Chemical shifts with the highest field signals of pyridine-d_5 as reference (7.19 ppm).

Table 4
^3J coupling constants measured on the saccharides of SAPO50 in 1:1 pyridine-d_5–D_2O

Residue	$^3J_{1,2}$	$^3J_{2,3}$	$^3J_{3,4}$	$^3J_{4,5eq}$	$^3J_{4,5ax}$	$^3J_{5ax,5eq}$	$^3J_{5,6}$
α-Rha*p* (**A**)	1.90 [a]	4.0 [a]	9.0 [b]		8.9 [b]		6.20 [a]
β-Fuc*p* (**B**)	8.11 [a]	8.5 [b]	3.0 [b]		3.0 [b]		6.20 [a]
β-Gal*p* (**C**)	7.63 [a]	9.1 [b]	3.0 [b]		4.1 [b]		*
β-Xyl *p* (**D**)	7.87 [a]	8.3 [b]	9.2 [b]	3.9 [b]	9.7 [b]	−11.0 [b]	
β-Xyl *p* (**E**)	7.63 [a]	8.3 [b]	9.5 [b]	5.1 [b]	10.0 [b]	−12.0 [b]	
β-Xyl *p* (**F**)	7.63 [a]	8.5 [b]	9.0 [b]	5.1 [b]	10.0 [b]	−12.0 [b]	
β-6-Deoxy-Glc*p* (**G**)	7.87 [a]	8.0 [b]	9.6 [b]		9.0 [b]		6.20 [a]
β-GlcA *p* (**H**)	7.39 [a]	8 [c]	*		*		

[a] Directly measured on the 1D spectrum (Hz ± 0.01 Hz).
[b] Measured on the subspectra obtained with the 1D-*z*-TOCSY sequence (Hz ± 0.1 Hz).
[c] Estimated from the H-1,H-2 correlation cross-peak in the high-resolution phase-sensitive DQF-COSY (Hz ± 1 Hz).
* Not observed.

of residue **B** shows only one correlation with **B**H-5 (unambiguously assigned by the DQF-COSY spectrum) whereas the anomeric proton shows correlation only with **B**H-2, **B**H-3, and **B**H-4, indicating the existence of a small $^{3}J_{H\text{-}4,H\text{-}5}$ value. This strongly supports a *galacto* configuration [39], and residue **B** was assigned to a β-fucopyranoside. This assignment was further confirmed by the values of the vicinal coupling constants obtained from 1D-*z*-TOCSY spectra. The β configuration is supported both by the values of $^{1}J_{CH\text{-}1}$ and $^{3}J_{H\text{-}1,H\text{-}2}$, and the presence of intense cross-peaks from **B**H-1 to **B**H-3 and **B**H-5 in the 1D-NOESY spectrum [40].

Residue C.—As for residue **B**, the cross-section of the TOCSY spectrum of residue **C** through the anomeric proton shows only three cross-peaks, thus indicating the existence of a small $^{3}J_{H\text{-}4,H\text{-}5}$ value and suggesting a *galacto* configuration. Moreover, the analysis of the 2D-HSQC-TOCSY spectra shows a correlation with a ^{13}C upfield shifted methylenic resonance at position 6 (the methylenic carbon was unambiguously identified based on the multiplicity-edited ge-PEP-HSQC experiment). Thus, residue **C** was assigned to a β-galactopyranoside. This assignment was further confirmed by the values of vicinal coupling constants and the intra-ring nOe network.

Residues D, E, and F.—The cross-sections of the TOCSY spectrum of these three residues through the anomeric protons show six signals, including H-1 and H-5 protons, and five carbons were observed in the HSQC-TOCSY spectra (τ_m 80 ms) in which the anomeric protons show connectivities with carbons C-2, C-3, C-4, and C-5. The methylenic nature of carbons C-5 was also unequivocally demonstrated by the negative sign of their one-bond proton–carbon correlation cross-peaks in the multiplicity-edited ge-PEP-HSQC. The pyranoside pattern, as the β configuration, was confirmed by intra-ring nOe correlations (H-1/H-3 and H-1/H-5ax) and vicinal proton–proton and direct proton–carbon coupling constant measurements. Thus, residues **D**, **E**, and **F** were assigned to β-xylopyranoside.

Residue G.—The anomeric proton of residue **G** shows correlations with methyl protons in the TOCSY spectrum. This indicates the presence of a large vicinal coupling among ring protons due to *trans*

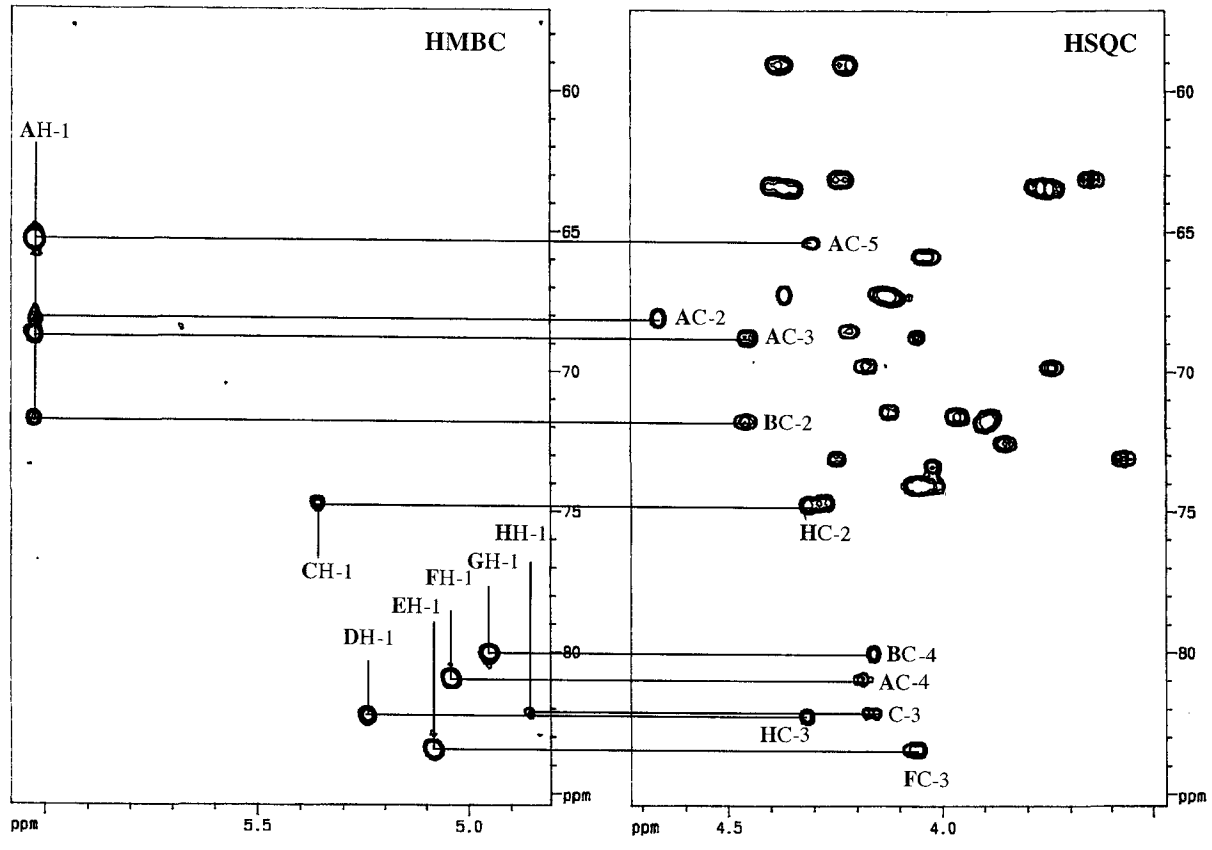


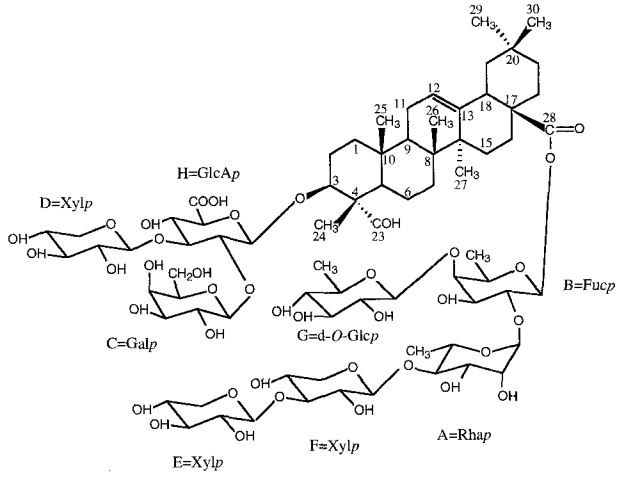
Fig. 5. Part of the HMBC spectrum (anomeric region in F_2) and of the multiplicity-edited HSQC spectrum (sugar ring resonances) recorded on SAPO50. The line shows the inter-residue long-range connectivities, from anomeric protons (HMBC) to carbons (HSQC).

diaxial orientations [38], suggesting a *gluco* configuration. This was confirmed by the precise measurement of the ^{3}J coupling constants in the 1D-*z*-TOCSY spectrum. Meanwhile, the $^{3}J_{H\text{-}1,H\text{-}2}$ value for residue **G** was 7.87 Hz, excluding the *rhamno* configuration and indicating a β-anomeric configuration supported by the $^{1}J_{CH}$ value of 157 Hz. Therefore, residue **G** was assigned to 6-deoxy-β-glucopyranoside. The procedure used to assign ^{13}C resonance was the same as above.

Residue H.—Difficulty was encountered for residue **H** due to signal overlap, posing a problem for the identification of the residue. Apparently the anomeric proton shows three cross-peaks, but inspection of their relative intensity and multiplicity on the cross-section (or better, on the 1D-*z*-TOCSY) shows the presence of a strongly coupled system. Only the **H**H-2 resonance could be unequivocally identified from the DQF-COSY spectrum, and only the assignment of **H**C-1 could be obtained by direct correlation with the HSQC spectrum. The complete assignment of ^{13}C and ^{1}H resonances was obtained from the 2D-HSQC-TOCSY experiment recorded with a small mixing time (16 ms). All three-bond correlations could be obtained from this single experiment, suggesting large vicinal coupling constants between the ring protons due to *trans* diaxial orientations. Thus, we presumed a *gluco* configuration for this residue. Additionally, a three-bond intra-ring correlation between **H**H-4 and a carboxylic carbon resonance (172.4 ppm) allowed us to assign this residue to a β-glucuronopyranosidic acid. The β configuration was deduced from the $^{3}J_{H\text{-}1,H\text{-}2}$ value (7.39 Hz) and by the $^{1}J_{CH}$ values (158 Hz).

Determination of linkages.—The final step in the elucidation of the structure represents the determination of the linkages among the sugar residues, as well as that between the carbohydrate chains and the triterpene. In principle, this information could be extracted from the 1D-NOESY experiments previously used to determine the configuration of each of the monosaccharides. In practice, however, a saponin of medium complexity such as SAPO50 (or even SAPO30), such a strategy leads to ambiguous results since, in several cases, nOes between each anomeric proton signal and the crucial methine or methylene proton signal of other saccharidic unit are severely overlapped. In this case, the ^{1}H–^{13}C long-range coupling through the glycosidic bond can be employed to resolve ambiguities. We have therefore used a gradient-enhanced HMBC experiment, which has the advantage of being a very sensitive method for establishing glycosidic linkages since HMBC (as HSQC and derivatives) is proton-detected. As the ^{13}C resonances of each sugar moiety have been unambiguously assigned previously, inter-residue correlations are easily identified. The inter-glycosidic long-range carbon correlations arising from the anomeric protons are shown in Fig. 5. Additionally, the large number of intra-ring cross-peaks (data not shown) observed in the HMBC spectrum support the ^{1}H and ^{13}C assignments discussed above. The anomeric proton of residue **D** is correlated with **H**C-3, and **C**H-1 is correlated with **H**C-2, identifying a trisaccharide segment. This sequence was also found in SAPO30. The trisaccharide **D–C–H** is linked to C-3 of the triterpene as indicated by a cross-peak between the anomeric proton of residue **H** and C-3 of the gypsogenin. The anomeric proton of residue **G** shows a correlation with **B**C-4, and **A**H-1 is correlated with **B**C-2. **E**H-1 is correlated to **F**C-3, and **F**H-1 to **A**C-4, leading to a pentasaccharide chain. The pentasaccharide segment **EFA–G–B** is linked to the ester group of the gypsogenin, as indicated by a strong connectivity between the anomeric proton of residue **B** and C-28 of the triterpene.

The strategy developed and used for SAPO50 demonstrates the structure elucidation of complex carbohydrates using high-field NMR. Modern techniques using pulse-field gradients are powerful tools for the assignment of ^{1}H and ^{13}C spectra. Such robust experiments are easy to implement for routine use, as careful adjustment of the different experimental parameters usually is not necessary. Moreover, despite the different anomeric configurations and ring forms for the eight sugars in SAPO50 (seven in SAPO30), complete resonance assignments and structure deter-



Scheme 1. Chemical structure of SAPO50.

mination were possible, based on the redundancy provided by the different NMR methods used. The chemical structure of SAPO50 is shown in Scheme 1. Note that SAPO30 only differs from SAPO60 by the nature of the saccharidic chain at position C-28 of the triterpene. Are the nature and the organization of the pentasaccharidic chain responsible for the haemolytic properties of SAPO50? This assumption seems to be supported by the fact that the monodesmoside, obtained by hydrolysis of the ester function and common to either SAPO50 or SAPO30, presents no haemolytic capabilities (unpublished results).

3. Experimental

Isolation of SAPO30 and SAPO50.—TLC (silica gel plate; 65:50:10 $CHCl_3$–MeOH–H_2O) analysis of the commercial Merck Saponin reveals that haemolytic constituents, identified by spraying blood reagent [41], have R_f values ranging from roughly 0.4 to 0.6. The low R_f non- or weakly haemolytic compounds were discarded by dialysis against water. Analysis of the sample after evaporation to dryness, shows a two-fold increase of the haemolytic activity with respect to the commercial Merck Saponin. This extract was subjected to C-18 column chromatography (60:40 MeOH–H_2O), and the fraction with the highest haemolytic activity was then isocratically chromatographed on a semi-preparative HPLC reverse-phase column, using 72:28 KH_2PO_4 (50 mM) buffer–MeCN as mobile phase. The two main products having retention time of 30 mn (SAPO30) and 50 mn (SAPO50) were isolated. Analysis of these two amorphous white powders resulted in a single spot on TLC and a single peak by analytical HPLC. The haemolytic activities of SAPO50 and SAPO30 show a 4- and a 100-fold decrease, respectively, with respect to the haemolytic activity of the commercial Merck Saponin.

General methods.—TLC was carried out on 0.25 mm Silica Gel 60 F_{254} E. Merck plates (20 cm × 20 cm). Thin-wall Visking bags (Polylabo) were used for dialysis (MW 6000–8000). Atmospheric pressure chromatography was carried out on a 50-cm length, 2.5-cm diameter glass column filled with 35 g of C-18 phase (Macherey Nagel). Analytical HPLC was performed on a Lichrospher 100 RP 18 (end-capped), 5-μm (4 mm × 250 mm i.d.) column (E. Merck). The flow rate was 0.7 mL min^{-1} (1750 psi), and the detection wavelength was 203 nm. To retain as far as possible the analytic separation conditions, the semi-preparative HPLC was performed on a similar column: Lichrospher 100 RP 18 (end-capped), 5 μm (10 mm × 250 mm i.d.) (E. Merck). The flow rate was 3 mL min^{-1} (2400 psi), and the detection wavelength was 203 nm. Nano-spin + 4000 D units (Gelman) were used for desalting, using ultra-centrifugation (5000 g). All chemical reagents and solvents are analytical or HPLC purity grade. The haemolytic activity was estimated using the haemolytic micromethod test previously described by Jurzysta [42].

Mass spectrometry.—Fast Atom Bombardment spectra (Xe, 8 keV, negative mode, matrix: thioglycerol) were obtained on a Jeol JMS.DX300 spectrometer. The fragmentation pattern obtained by negative FABMS for SAPO30 was similar to previously reported data [8]. The fragmentation pattern obtained for SAPO50 confirms our NMR structure determination: m/z 1683 $[M - H + K]^-$, 1551 $[M - H + K - 132]^-$, 1419 $[M - H + K - 132 - 132]^-$, 1213 $[M - H + K - (162 + 176 + 132)]^-$. In particular, they indicate that there are no non-protonated groups (e.g. sulfate) attached to one or more hydroxyl groups of the sugars.

NMR spectroscopy.—Samples (10 mg) were dissolved in 1:1 pyridine-d_5–D_2O (0.4 mL). Spectra were recorded on a Bruker AMX-600 spectrometer equipped with a z-gradient triple-resonance (^1H, ^{15}N, ^{13}C) probe. The upfield peak of the three pyridin signals in ^1H and ^{13}C spectra were taken as the internal standard (δ 7.19 ppm for ^1H and δ 123.5 ppm for ^{13}C). All experiments were carried out at 47 °C for SAPO50 (43 °C for SAPO30). The following techniques and parameters were used.

All proton spectra were recorded with a spectral window of 9000 Hz (15 ppm). Selective 1D experiments were performed using the regular DANTE-Z sequence for frequency selection. The parameters used for this sequence are listed in the legend of Fig. 4. In the 1D-NOESY experiment, the mixing time was 200 ms. For each experiment, 512 scans (4 dummy scans) were acquired (32K complex points), with a repetition time of about 3 s. The WALTZ-16 [43] sequence was used for isotropic mixing (250 ms) in the 1D-z-TOCSY experiment, at a field strength of 5 kHz (corresponding to a π pulse of 100 μs). This spin-lock field is sufficient to cover the limited frequency range of the carbohydrate proton resonances, and excellent spin-locking conditions were obtained without noticeable heating of the sample (chemical shifts did not change more than 0.01 ppm compared to the 1D spectrum). For each experiment, 128 scans

(4 dummy scans) were acquired (32K complex points), with a repetition time of about 3 s. The time-domain matrix size of 2D-z-TOCSY and 2D-DQFCOSY was 4K × 512 ($F_2 \times F_1$) complex points. Sixteen scans were co-added for each t_1 increment, with a repetition rate of 1.5 s. Before processing, the time-domain matrices were zero-filled in F_1 to 2K × 1K real points. Squared-cosine-bell and cosine-bell apodization functions were applied on the F_2 and F_1 time-domain, respectively, before Fourier transform. As for the 1D-z-TOCSY, the WALTZ-16 sequence was used for isotropic mixing (150 ms) in the 2D-TOCSY experiment.

Proton-detected 2D heteronuclear spectra were generally recorded with spectral widths of 4200 Hz (7 ppm) in F_2 and 22,500 Hz (150 ppm) in F_1, meaning that the signal of the aldehydic group from gypsogenin and the pyridin residual signals were folded in both the proton and the carbon dimensions in order to gain digital resolution. The time-domain matrix sizes were 2K × 512 complex points, and were zero-filled in F_1 to 1K × 1K real points before processing. The same apodization functions were applied as in 2D homonuclear experiments before Fourier transform. ^{13}C Decoupling during the entire acquisition period was done by composite pulse decoupling using the GARP [44] sequence. We are indebted to Dr. Teodor Parella who supplied us with the PEP version (Preservation of Equivalent Pathway) of the sequence of multiplicity-edited ge-PEP-HSQC before publishing [3]. HSQC-TOCSY and ge-HMBC were used as previously described in the literature, and z-spoils were implemented as pulse-field gradients in the HSQC-TOCSY sequence as discussed in the present paper. As for homonuclear 2D-TOCSY, the WALTZ-16 sequence was used for isotropic mixing (16 and 80 ms) in the HSQC-TOCSY, with a similar spin-locking RF field. Additionally, an unfolded HMBC experiment was recorded (^1H: 15 ppm, ^{13}C: 300 ppm) in order to reach the long-range correlations with down-field shifted groups (carboxylic, esters, etc.). On the other hand, a high-^1H-resolution (4K complex points), highly ^{13}C-folded (50 ppm, 128 points) HSQC was recorded without ^{13}C decoupling during the acquisition in order to measure accurately the anomeric $^1J_{CH}$ coupling constants. Sixteen scans were added for each t_1 increment in the HSQC experiment, 32 for the HSQC-TOCSY and the ge-HMBC experiments, with a repetition rate of 1.5 s.

In DQF-COSY, z-TOCSY, and HSQC-TOCSY experiments, quadrature detection in the F_1 dimension was obtained by the States-TPPI method [45]. The multiplicity-edited ge-PEP-HSQC experiment was acquired with the N and P experiments interleaved, and processing was applied as described by Palmer et al. [16] (echo–antiecho mode in the Bruker software). The ge-HMBC experiment was recorded with phase modulation (magnitude mode). Data were processed in an off-line station using indifferently the UXNMR Bruker software, or the Gifa (version 4) software [46]. All pulse-sequence micro-programs (Bruker AMX version) used in the present paper are available with extensive parameter description upon request to the senior author.

[3] The pulse program is now available in the electronic-mail NMR pulse-sequence library at http://rmn3.uab.es/enciclo.html.

Acknowledgements

We thank the company ABX (rue du Caducée, Parc Euromédecine, F-34184 Montpellier) for financial support.

References

[1] K. Hostettmann and A. Martson, in *Chemistry and Pharmacology of Natural Products: Saponins*, (Eds.) Cambridge University Press, 1995 pp 232–304.
[2] R. Tschesche and G. Wulff, *Fortschr. Chem. Org. Naturst.*, 30 (1973) 461–606.
[3] S.K. Agarwal and R.P. Rastogi, *Phytochemistry*, 13 (1974) 2623–2645.
[4] Y.C. Kim, R. Higuchi, and T. Komori, *Liebigs Ann. Chem.*, (1992) 941–946.
[5] K. Tori, S. Seo, Y. Yoshimura, M. Nakamura, Y. Tomita, and H. Ishii, *Tetrahedron Lett.*, 46 (1976) 4167–4170.
[6] A. Penders and C. Delaude, *Carbohydr. Res.*, 263 (1994) 79–88.
[7] G. Massiot, C. Lavaud, D. Guillaume, L. Le Men-Olivier, and G. Van Binst, *J. Chem. Soc., Chem. Commun.*, 728 (1986) 1485–1487.
[8] D. Frechet, B. Christ, B. Monegier du Sorbier, H. Fischer, and M. Vuilhorgne, *Phytochemistry*, 30 (1991) 927–931.
[9] L. Ma, G. Tu, S. Chen, R. Zhang, L. Lai, X. Xu, and Y. Tang, *Carbohydr. Res.*, 281 (1996) 35–46.
[10] N.H. Jacobsen, W.J. Fairbrother, C.R. Kensil, A. Lim, D.A. Wheeler, and M.F. Powell, *Carbohydr. Res.*, 280 (1996) 1–14.

[11] J.R. Garbow, D. Weitekamp, and A. Pines, *Chem. Phys. Lett.*, 93 (1982) 504–509.
[12] T. Parella, *Magn. Reson. Chem.*, 34 (1996) 329–347.
[13] W. Wilker, D. Leibfritz, R. Kesserbaum, and W. Bermel, *Magn. Reson. Chem.*, 31 (1993) 287–292.
[14] J. Keeler, R.T. Clowes, A.L. Davies, and E.D. Laue, *Methods Enzymol.*, 239 (1993) 145–207.
[15] J. Cavanagh and M. Rance, *Annu. Rep. NMR Spectrosc.*, 27 (1993) 1–58.
[16] A.G. Palmer, J. Cavanagh, P.E. Wright, and M. Rance, *J. Magn. Reson.*, 93 (1991) 151–170.
[17] G. Kontaxis, J. Stonehouse, E.D. Laue, and J. Keeler, *J. Magn. Reson. A*, 111 (1994) 70–76.
[18] J.A. Gavin, J.L. Pons, and M.A. Delsuc, *J. Magn. Reson. A*, 122 (1996) 64–66.
[19] T. Parella, F. Sanchez-Ferrando, and A. Virgili, *J. Magn. Reson. A*, 117 (1995) 78–83.
[20] F.J. Moy, H.A. Sheraga, J.F. Liu, R. Wu, and G. Montelione, *Proc. Natl. Acad. Sci. USA*, 86 (1989) 9836–9840.
[21] G. Bodenhausen and D.J. Ruben, *Chem. Phys. Lett.*, 69 (1980) 185–189.
[22] L. Lerner and A. Bax, *J. Magn. Reson.*, 69 (1986) 375–380.
[23] M. Rance, O.W. Sørensen, G. Bodenhausen, R.R. Ernst, and K. Wüthrich, *Biochem. Biophys. Res. Commun.*, 117 (1983) 479–495.
[24] A. Bax and M.F. Summers, *J. Am. Chem. Soc.*, 108 (1986) 2093–2094.
[25] A. Bax and D. Marion, *J. Magn. Reson.*, 78 (1988) 186–191.
[26] P.L. Rinaldi and P.A. Keifer, *J. Magn. Reson. A*, 108 (1994) 259–262.
[27] F.W. Wehrly and T. Nichida, *Prog. Chem. Org. Nat. Compd.*, 36 (1979) 24–122.
[28] L. Braunschweiler and R.R. Ernst, *J. Magn. Reson.*, 53 (1983) 521–528.
[29] D.G. Davis and A. Bax, *J. Am. Chem. Soc.*, 107 (1985) 2820–2821.
[30] M. Rance, *J. Magn. Reson.*, 74 (1987) 557–564.
[31] P. Berthault, F. Djedaïni, and B. Perly, *J. Magn. Reson.*, 91 (1991) 102–112.
[32] D. Boudot, C. Roumestand, F. Toma, and D. Canet, *J. Magn. Reson.*, 90 (1990) 221–227.
[33] O.W. Sørensen, M. Rance, and R.R. Ernst, *J. Magn. Reson.*, 56 (1984) 527–534.
[34] S. Subramanian and A. Bax, *J. Magn. Reson.*, 71 (1987) 325–330.
[35] D. Boudot, D. Canet, J. Brondeau, and J.C. Boubel, *J. Magn. Reson.*, 83 (1989) 428–439.
[36] C. Roumestand, D. Canet, N. Mahieu, and F. Toma, *J. Magn. Reson. A*, 105 (1994) 168–181.
[37] R. Kasai, M. Okihara, J. Asakawa, K. Mizutani, and O. Tanaka, *Tetrahedron*, 35 (1979) 1427–1432.
[38] C. Altona and C.A.G. Haasnoot, *Org. Magn. Reson.*, 40 (1980) 417–429.
[39] C. Abeygunawardana, C.A. Bush, and J.O. Cisar, *Biochemistry*, 29 (1991) 234–248.
[40] S.W. Homans, *Prog. Nucl. Magn. Reson. Spectrosc.*, 22 (1990) 55–81.
[41] E. Stahl, E. Dumont, H. Jork, L. Kraus, K.E. Rozumek, and P.J. Schorn, in *Analyse Chromatographique et Microscopique des Drogues*, (Eds) Entreprise moderne d'Edition et Technique de Documentation, Paris, 1975 p 234.
[42] M. Jurzysta, *Acta Agrobot.*, 32 (1979) 5–11.
[43] A.J. Shaka, J. Keeler, and R. Freeman, *J. Magn. Reson.*, 53 (1983) 313–334.
[44] A.J. Shaka, P.B. Barker, and R. Freeman, *J. Magn. Reson.*, 64 (1985) 547–552.
[45] D. Marion, M. Ikura, R. Tschudin, and A. Bax, *J. Magn. Reson.*, 85 (1989) 393–399.
[46] J.L. Pons, T.E. Malliavin, and M.A. Delsuc, *J. Biol. NMR*, 8 (1996) 445–452.