Structural Elucidation of Alfalfa Root Saponins by Mass Spectrometry and Nuclear Magnetic Resonance Analysis

Georges Massiot,* Catherine Lavaud, and Louisette Le Men-Olivier
Faculté de Pharmacie, UA au CNRS n° 492, 51, rue Cognacq-Jay, 51096 Reims Cédex, France
Georges Van Binst
Vrije Universiteit Brussels, Fakulteit der Wetenschappen, Pleinlaan 2, B-1050, Brussel, Belgium
Stephen P. F. Miller and Henry M. Fales
National Institutes of Health, Bethesda, Maryland 20892, U.S.A.

Structures of saponins may be fully elucidated using a combination of n.m.r. and m.s. instead of the usual degradation techniques. Well resolved proton n.m.r. spectra and chromatographic separation can be obtained from peracetylated mixtures of saponins. Under these conditions, genins are identified mainly by ^{13}C n.m.r. spectroscopy; sugars and their points of attachment are determined by examining ^{1}H chemical shifts and coupling constants following assignments made by 2D COSY and relayed COSY experiments. Sugar chains are sequenced by observing intersugar connectivities through long-range couplings (delayed COSY) or n.O.e. Alternatively, the use of Californium plasma desorption m.s. provides quasimolecular ions and fragments corresponding to the sequential rupture of the chain of sugars. Free carboxylic acids are located on an acetylated saponin by looking at the disappearance, in the presence of Eu complexes, of the carbon situated α to the carbonyl in the ^{13}C n.m.r. spectrum. The above methods have been employed to determine the structures of the seven major saponins of alfalfa root, which are allegedly responsible for the antifeedant properties of the vegetable.

Structural determination of saponins requires identification of their building blocks (genins, sugars, acids) as well as of their linkages. This problem has been addressed by Tschesche in an article which reviews a set of solutions mainly based on selective degradation.[1] Non-degradative methods such as mass spectrometry[2-4] or ^{13}C n.m.r. spectroscopy[5-7] offer new possibilities in the field. With the increasing resolution of modern spectrometers, ^{1}H n.m.r. spectroscopy now allows more detailed investigation of the structures of these complex molecules.

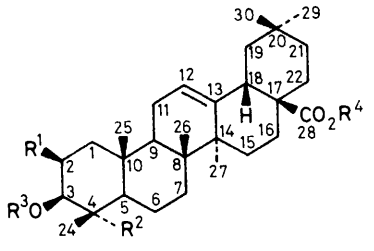
We report the use of n.m.r. and m.s. in the structural elucidation of seven major saponins of alfalfa roots, whose relevance to nutritional problems is well known. The saponins investigated are 28-*O*-β-D-glucopyranosyl-3-*O*-β-D-glucopyranosylmedicagenate (**1**; R = H), 3-*O*-[β-D-galactopyranosyl-(1 ⟶ 2)-α-L-arabinopyranosyl]-28-*O*-β-D-glucopyranosylhederagenin (**2**; R = H), 23- or 28-*O*-β-D-glucopyranosyl 3-*O*-[β-D-galactopyranosyl-(1 ⟶ 2)-β-D-glucopyranosyl]-medicagenate (**3**; R = H), 3-*O*-[α-L-arabinopyranosyl(1 ⟶ 2)-β-D-glucopyranosyl-(1 ⟶ 2)-α-L-arabinopyranosyl]-28-*O*-β-D-glucopyranosylhederagenin (**4**; R = H), 28-*O*-β-D-glucopyranosyl 3-*O*-[α-L-rhamnopyranosyl-(1 ⟶ 2)-β-D-glucopyranosyl-(1 ⟶ 2)-β-D-glucopyranosyl]-medicagenate (**5**; R = H), 28-*O*-[β-D-xylopyranosyl-(1 ⟶ 4)-α-L-rhamnopyranosyl-(1 ⟶ 2)-α-L-arabinopyranosyl]-3-*O*-β-D-glucopyranosylmedicagenate (**6**; R = H), and 28-*O*-[β-D-xylopyranosyl-(1 ⟶ 4)-α-L-rhamnopyranosyl-(1 ⟶ 2)-α-L-arabinopyranosyl] 3-*O*-[β-D-glucopyranosyl-(1 ⟶ 2)-β-D-glucopyranosyl]-medicagenate (**7**; R = H).

The saponins were extracted and purified in a standard fashion. Although some purification was obtained on underivatized material, more success was met after full acetylation [Ac_2O, 4-(dimethylamino)pyridine (DMAP)] followed by repeated conventional column chromatography and preparative t.l.c. (p.l.c.). All the n.m.r. investigations reported here have been performed on the acetylated derivatives. The reconversion into underivatized material for biological assays was performed by mild alkaline hydrolysis. Genins [hederagenin (**8**) and medicagenic acid (**9**)], were identified by acidic microhydrolysis and t.l.c. detection. Each individual saponin† was examined by n.m.r. and genins were also identified by their ^{1}H and ^{13}C n.m.r. spectra. In the ^{13}C n.m.r. spectra (Table 1) all the signals of the triterpenoids were unchanged with regard to the parent compounds, except for the points of substitution: C-3 (10—12 p.p.m. downfield) and C-28 (3 p.p.m. upfield). More interestingly, the ^{1}H signals for the methyl groups were easily recognizable and hederagenin derivatives (**2**) and (**4**) showed five methyls at a field higher than 1 p.p.m. (*i.e.* $\delta < 1$) while medicagenic acid derivatives (**1**), (**3**), and (**5**)—(**7**) had only three methyl singlets in this area. Observation of a doublet for 3-H in these latter cases was another method for detection of medicagenic acid derivatives.

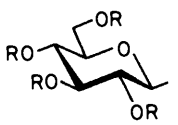
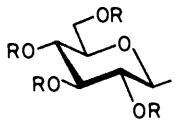
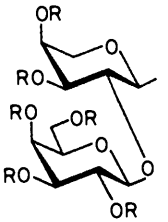
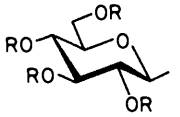
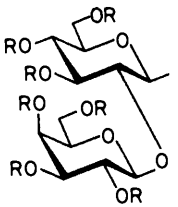
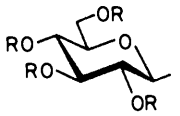
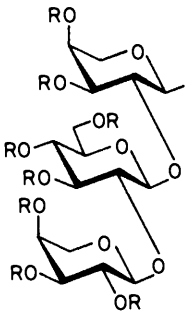
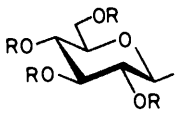
Identification of the Sugars and of their Branching Points.—After acetylation of the saponins, the sugar protons showed signals in the δ 3—5.5 range and their assignment is the basis of the sugar identification process. Increasing length of the sugar chains generated an entangled CHO area whose analysis was performed by using 2D correlated spectroscopy (COSY). An ordinary COSY experiment was first run on the whole spectrum to establish gross connectivities and, in a second COSY experiment, digital resolution was enhanced by selecting a window containing only the sugar resonances. Quality of the spectra was improved by using a double quantum filter to decrease the size of the acetate singlets which might otherwise fold over in the sugar proton area[8] or by showing the correlation map in the phase sensitive mode.[9] Such experiments generally allowed the construction of unambiguous correlation maps, and, since in normal sugars the coupling path is not interrupted by quaternary carbons, it is feasible to follow a route from 1-H to 5-H in pentoses and 6-H in hexoses. Figure 1 shows such a map obtained for the four-sugar saponin (**4**).

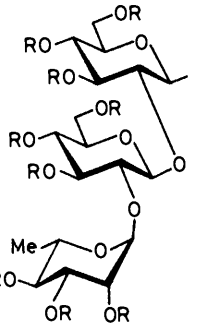
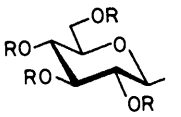
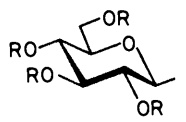
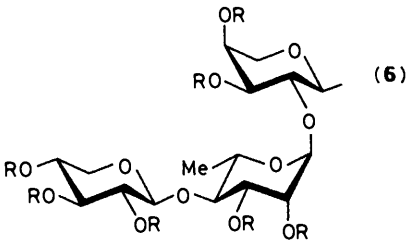
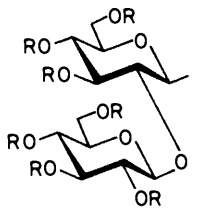
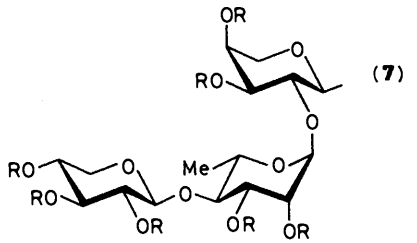
Starting from the anomeric proton signals, one can extract two seven-spin systems of hexoses and two six-spin systems of pentoses. The shape of the cross peaks of the hexoses indicates a value of *ca.* 8 Hz for the vicinal coupling constants with the

† From now on, the term saponin and numbers (**1**)—(7) will refer to the fully derivatized material.



$R^1 = R^3 = R^4 = H$, $R^2 = CH_2OH$ **(8)**
$R^1 = OH$, $R^2 = CO_2H$, $R^3 = R^4 = H$ **(9)**
$R^1 = OH$, $R^2 = CH_2OH$, $R^3 = R^4 = H$ **(10)**

R^1	R^2	R^3	R^4	
OH	CO_2H			**(1)**
H	CH_2OR			**(2)**
OH	CO_2(H or R^4)		 or H	**(3)**
H	CH_2OR			**(4)**

R^1	R^2	R^3	R^4	
OH	CO_2H			**(5)**
OH	CO_2H			**(6)**
OH	CO_2H			**(7)**

R = H or Ac

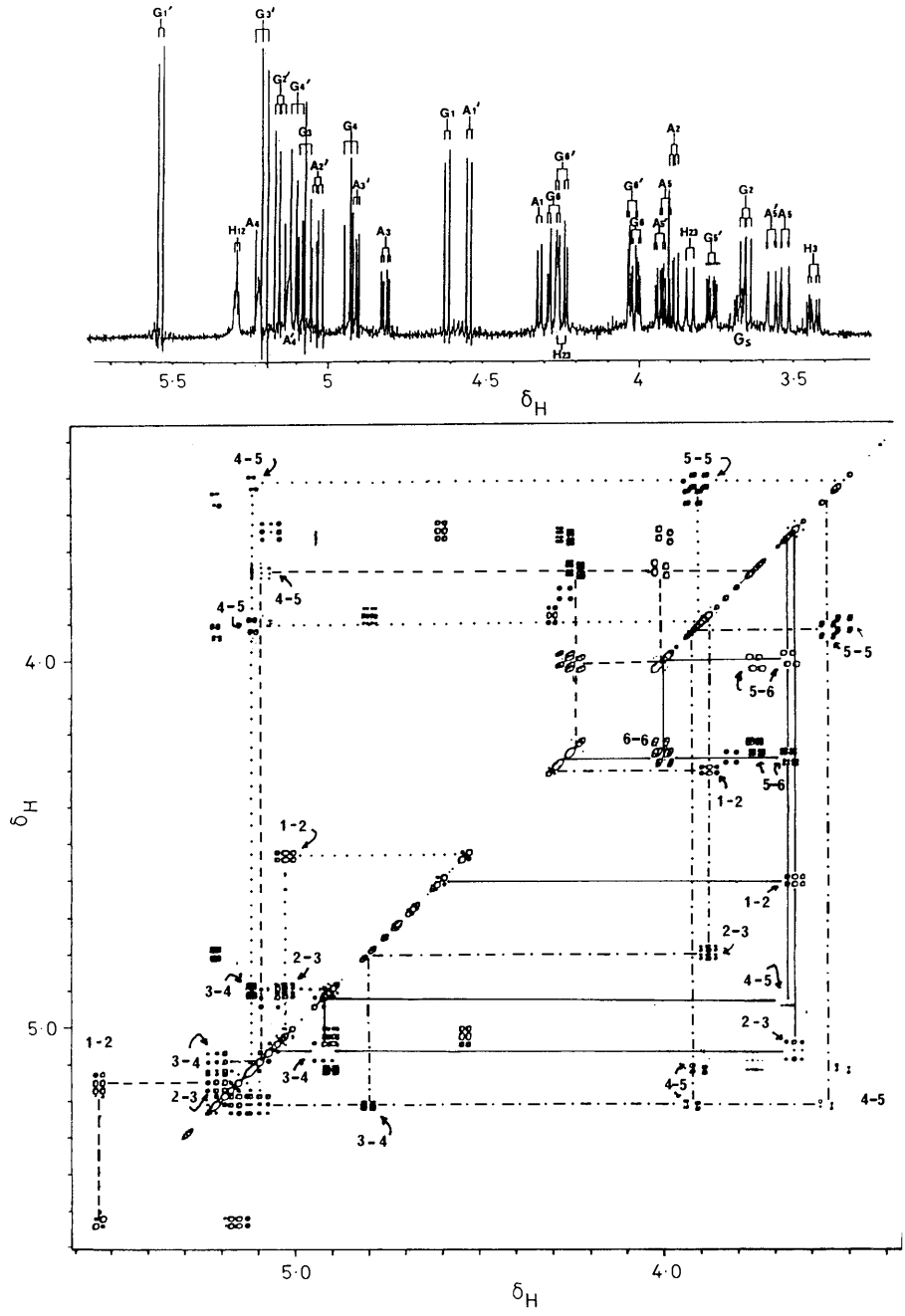


Figure 1. A 400 MHz two-dimensional COSY-45 spectrum of saponin (**4**). A 2.25 p.p.m. window was selected for acquisition in the F_2 domain. Parameters: F_2 = 1 K data points; F_1 = 512 W; relaxation delay 350 ms; pulse widths 14 μs (90°) and 7 μs (45° mixing pulse); initial delay 5 μs; increment 1.11 ms; number of scans 16. Transformed with a squared sine bell and shown symmetrized. Reference spectrum at the top is the fully assigned 500 MHz spectrum of compound (**4**). Inner glucose correlations ———, terminal glucose correlations -----, arabinose correlations –·–·–, terminal arabinose correlations ·····

exception of $J_{5,6}$ which is close to 3 Hz. These values indicate that all the acetate groups are equatorial and therefore that the hexoses are glucoses. In the pentoses, one observes that all couplings are large except those with 4-H; the pentose whose 4-OH is axial is arabinose.

The four sugars of saponin (**4**) are thus two glucoses and two arabinoses; given previous work on a alfalfa saponin sugars, it seems safe to assume that the sugars are D-glucose and L-arabinose even though this cannot be proven by n.m.r. spectroscopy. The values of $J_{1,2}$ for all these sugars indicated that they are β-D-glucoses and α-L-arabinoses.

It was clear from the chemical shifts that C*H*OAc resonances are normally found as $\delta > 4.5$ whereas CH_2OAc resonances are at $\delta \sim 4$. Observation of 5-H of glucoses and arabinoses at $\delta \sim 4$ indicated that the C-5 of these sugar was engaged in an ether linkage, *i.e.* that the sugars are in the pyranose form. The abnormally high shielding of the 2-H protons of two of the sugars showed that these positons were not acetylated; it is therefore concluded that these are branching points for the sugar chain. The two other sugars were fully acetylated; hence they must be terminal.

The final information which can be extracted from these spectra concerned the chemical shifts of the anomeric protons; three of them were found at δ 4.4 ± 0.2 and corresponded to

ethers. The δ 5.54 doublet of 1-H of one glucose is best explained by substituting this sugar on the carboxylic acid function of hederagenin. Since this is a terminal sugar, the other sugar chain must consist of three units with a terminal arabinose, which it is not possible to sequence using chemical-shift arguments. This chain of sugars was placed on C-3 of hederagenin since 3-H appeared at higher field than the resonances corresponding to the C-23 methylene protons. A similar analysis was performed on saponin (**2**), which contains hederagenin, arabinose, glucose, and galactose which are distinguished by the shape of their 4-H–5-H correlations.

In the COSY spectrum of saponin (**6**) three correlations were found between the low- and high-field parts of the spectrum; they are the 12-H—11-H and 2-H—1-H correlations and also a correlation of a methyl doublet with a C*H*O proton (δ 3.8). Starting from the methyl doublet, it was possible to identify a spin-system corresponding to a rhamnosyl residue substituted on its C-4. Other sugars were an arabinose substituting one of the acid functions of medicagenic acid and branched on its C-2 and a terminal glucose. The fourth sugar is a pentose whose vicinal coupling constants are all larger than 7 Hz; it was identified as a terminal xylose. At this point, as for saponin (**4**), it was not possible fully to elucidate the structure of compound (**6**); it remained to sequence the chain of sugars and to decide on which carboxylic acid to place it. These points will be dealt with later. Complete spin–spin analysis was similarly performed for the seven major saponins of alfalfa roots. Results are summarized in Table 2.

In cases where overlap prevented unambiguous assignments to be made, a very useful modification of the COSY was the relayed COSY in which relayed connectivities appeared as well as direct connectivities.[10] Observation, for example, of $J_{1,2}$ and of 'relayed' $J_{1,3}$ within the sugar residues allowed accurate investigations in the δ4.5—5.2 range. In practice, with only two well separated proton signals, one can identify the resonances of all the protons of a hexose.

Sequence of the Sugar Chain using ^{1}H *N.m.r. Techniques.*— Among the compounds isolated in this study, the two- and three-sugar saponins (**1**), (**2**), and (**3**) do not require a sequencing of their sugar chain since these are limited to two units. Four-sugar saponin (**5**) possesses two inner glucoses similarly substituted on their C-2 positions and terminal glucose and rhamnose; the 'degeneracy' of the inner residues again makes sequencing unnecessary. Problems arise with saponins (**4**), (**6**), and (**7**) where two or three formulae may be written for each compound.

Observation of interproton couplings through the glycosidic bonds would be a solution to the problem and although these ^4J couplings are not usually resolved in 1D-spectra, they can be detected in delayed-COSY experiments using long delays before acquisition, typically 400 ms.[11] When applied to saponin (**6**), the experiment showed a large number of correlations, the majority of which were residual 'normal' cross peaks. Among new modulations, however, the most interesting ones were those related to the anomeric protons (Figure 2). Some structurally important long-range couplings were thus observed between rhamnose 1-H and arabinose 2-H and between xylose 1-H and rhamnose 4-H; these correlations allowed sequencing of the sugar chain as Xyl-(1 ⟶ 4)-Rha-(1 ⟶ 2)-Ara. Other less significant correlations are found between xylose 1-H and 5-H, rhamnose 1 ⟶ 5 and 4 ⟶ 6. The same experiment, when run on saponin (**4**), yielded correlations between inner glucose 1-H and inner arabinose 2-H and between inner glucose 2-H and terminal arabinose 1-H. In these examples no coupling is

Table 1. ^{13}C N.m.r. chemical shifts. Spectra of compounds (**1**) and (**4**)—(**9**) were assigned by 2D C–H correlation n.m.r. spectroscopy

	Structure								
Carbon	(**1**)	(**2**)	(**3**)	(**4**)	(**5**)	(**6**)	(**7**)	(**8**)	(**9**)
C-1	42.8	38.2	42.9	38.1	43.0	43.5	42.9	38.1	44.0
C-2	69.4	25.2	71.7	25.3	70.2	69.8	70.7	26.4	71.2
C-3	85.8	83.8	87.4	83.9	87.1	86.0	86.9	76.4	75.2
C-4	52.3	41.8	52.1	41.5	52.3	52.6	52.0	41.7	53.4
C-5	51.8	47.7	51.7	47.7	52.0	52.5	51.8	49.7	52.2
C-6	20.5	18.0	20.2	17.8	20.1	20.6	20.1	18.5	21.1
C-7	31.8	32.0	31.7	31.8	31.8	32.8	32.2	32.4	32.4
C-8	39.6	39.4	39.6	39.3	39.7	40.1	39.5	39.3	39.8
C-9	48.1	47.8	48.0	47.7	48.1	48.5	49.1	47.5	48.3
C-10	36.2	36.6	36.2	36.5	36.3	36.4	36.2	36.9	36.4
C-11	22.7	22.9	22.7	22.8	22.8	23.4	23.0	23.1	23.1
C-12	122.7	122.9	122.6	127.7	122.6	122.5	122.1	122.2	122.3
C-13	142.8	143.1	142.9	142.8	142.9	143.5	143.3	143.6	144.0
C-14	41.8	41.7	41.8	41.9	41.8	42.3	41.8	41.7	42.0
C-15	27.5	27.7	27.5	27.6	27.6	27.8	27.3	27.7	27.7
C-16	23.1	23.5	23.4	23.4	23.4	23.8	23.4	23.4	23.6
C-17	46.8	46.8	46.7	46.7	46.8	47.2	47.9	46.7	46.8
C-18	41.2	41.2	40.9	41.0	41.0	41.5	41.1	41.3	41.4
C-19	45.7	45.8	45.7	45.7	45.8	46.4	45.9	45.9	46.0
C-20	30.6	30.6	30.6	30.5	30.6	31.0	30.6	30.7	30.7
C-21	33.7	33.8	33.7	33.7	33.7	34.2	33.7	33.9	34.0
C-22	32.4	32.7	32.3	32.6	32.5	32.5	31.9	32.4	32.4
C-23	181.0	65.2	175.6*	65.5	179.0	180.5	178.6	71.3	178.5*
C-24	12.8	12.7	12.7	12.5	12.8	13.2	12.7	11.6	12.4
C-25	16.0	15.8	16.5	15.7	16.8	16.4	16.5	15.7	16.6
C-26	16.7	17.1	16.8	16.9	16.9	17.0	16.6	16.9	16.8
C-27	25.7	25.6	25.7	25.5	25.3	26.0	25.7	26.0	26.1
C-28	175.6	175.8	178.4*	175.5	175.5	175.8	175.5	178.2	178.4*
C-29	32.9	33.0	33.0	32.9	33.0	33.1	33.0	33.1	33.1
C-30	23.4	23.6	23.4	23.4	23.4	23.9	23.6	23.6	23.6

Table 1 (*continued*)

Carbon	Structure (1)	(2)	(3)	(4)	(5)	(6)	(7)
	β-D-glucose	α-L-arabinose	β-D-glucose	α-L-arabinose	β-D-glucose	β-D-glucose	β-D-glucose
C-1	101.4	103.3	103.2	103.8	103.2	101.3	102.8
C-2	69.9	76.4	74.8	72.5	73.0	72.4	74.4
C-3	72.3	72.7	71.2	72.7	74.6	72.7	69.6
C-4	71.1	68.2	68.7	68.1	68.7	68.8	68.2
C-5	71.8	62.9	72.0	62.7	71.5	72.1	71.6
C-6	61.4		61.4		61.3	61.8	61.8
	β-D-glucose	β-D-galactose	β-D-galactose	β-D-glucose	β-D-glucose	α-L-arabinose	β-D-glucose
C-1	91.5	103.3	99.6	101.8	99.6	92.8	99.6
C-2	67.9	71.3	70.2	78.7	78.6	74.5	71.1
C-3	72.8	71.9	72.8	75.0	73.9	72.4	72.1
C-4	68.4	68.6	68.4	69.0	68.7	66.8	68.0
C-5	72.4	75.1	72.4	71.3	71.8	62.7	71.9
C-6	61.8	62.3	61.9	62.3	61.9		61.7
		β-D-glucose	β-D-glucose	α-L-arabinose	α-L-rhamnose	α-L-rhamnose	α-L-arabinose
C-1		91.7	91.6	101.6	98.5	98.2	91.8
C-2		70.1	69.8	69.6	69.8	71.4	73.5
C-3		73.0	72.5	70.8	68.1	70.1	74.7
C-4		68.0	67.9	68.0	70.5	76.9	66.2
C-5		72.6	72.5	63.5	67.4	67.9	61.0
C-6		61.7	61.7		16.9	17.9	
				β-D-glucose	β-D-glucose	β-D-xylose	α-L-rhamnose
C-1				91.6	91.6	101.8	97.7
C-2				70.1	69.8	71.3	69.9
C-3				72.9	72.7	71.3	72.5
C-4				67.9	67.8	69.5	76.2
C-5				72.5	72.4	62.2	67.7
C-6				61.6	61.3	—	17.9
							β-D-xylose
C-1							100.9
C-2							70.9
C-3							70.7
C-4							69.0
C-5							62.4
C-6							

* Assignments may be interchanged within a column.

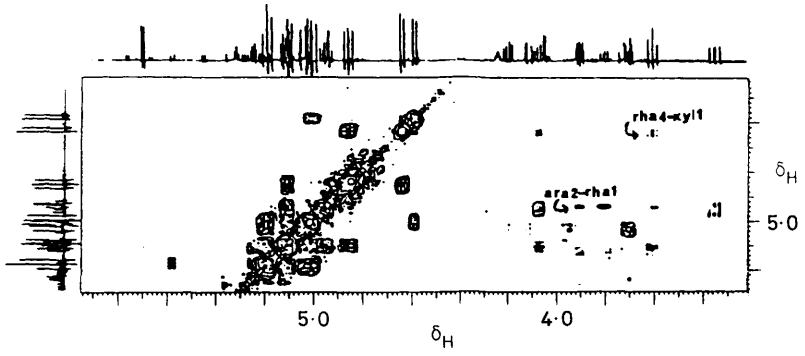


Figure 2. A 500 MHz delayed COSY spectrum of saponin (**6**) (sugar part only is displayed). Parameters: spectral width 5.8 p.p.m.; F_2 = 2 K data points; F_1 = 512 W; zero-filling prior to transformation; relaxation delay 1.5 ms; pulse widths 18 μs (90°), 18 μs; initial delay 3 μs; increment 0.346 ms; evolution delay 400 ms. Transformed with unshifted sine bell and symmetrized

found between 3-H of the genin and the anomeric proton of the first sugar in the chain. The structure of compound (**7**), which is related to saponin (**6**), was established by comparison of their ^1H and ^{13}C chemical shifts and the sequence was further confirmed by mass spectrometry (*vide infra*).

As an alternative to the delayed COSY approach it has been proposed[12] that n.O.e. difference spectroscopy could be used to establish intersugar linkages. In contrast with the previous report, we found difficulty in observing n.O.e.s in $CDCl_3$ at 500 MHz because of the lack of viscosity of the medium and of an

unfavourable $\omega\tau_c$ factor. More intense and more reproducible correlations were obtained when the experiment was run in [$^2H_6$] DMSO; the change of solvent induces dramatic and non-predictable shifts, which must be controlled by running a COSY experiment in dimethyl sulphoxide (DMSO). The precision and the importance of the NOESY experiment was demonstrated by the correlations found between the arabinose anomeric protons (δ 4.42 and 4.59) and two almost synchronous protons (δ 3.45). One of these correlations relates arabinose 1-H to proton 3-H of the genin, while the second one is due to the other arabinose 1-H and the inner glucose 2-H. Other important NOESY modulations were found between 1—3 diaxially oriented protons and between 12-H and 18-H.

Sequencing of Sugars by Mass Spectrometry.—Quasi-molecular ions and fragment ions may be obtained from underivatized glycosides and even complex saponins by a variety of 'soft' ionization techniques including field desorption, fast atom or ion bombardment, chemical ionization, and laser methods.[13] On the other hand, little corresponding data exist on acetylated saponins other than the simple glycosides.

Californium plasma desorption m.s. (PDMS) has been shown to be useful in highly polar, high molecular weight samples[14] and its extension to these acetylated saponins appeared feasible. In all cases quasimolecular ions (M·Na^+) were obtained; because the time-of-flight system does not permit resolution of isotopes in the quasimolecular ion cluster, elemental compositions were calculated using isotopic average atomic weight (*e.g.* C 12.011 15). An accuracy of 1—2 mass units is obtained in this mass range, allowing determination of the molecular composition given previous knowledge of the genin and sugars.

Quasimolecular ions were accompanied by more intense fragments which arose from sequential ruptures of the glycosidic bonds. Following n.m.r. or hydrolytic identification of the sugars, these fragmentations also allowed sequencing to occur (Table 3). These results with acetylated saponins were in direct contrast to those obtained in a study of non-acetylated glycosides related to digitonin,[15] where M·Na^+ ions were relatively more intense and the sugar acylium ions were barely detectable. Part of the difference in M·Na^+ ion intensity was traced to the significantly greater concentration of sodium salt residues in the unprotected saponin samples (presumably due to leaching from glass by alcoholic solvents during sample preparation). On the other hand, the much greater abundance of sugar acylium ions in the spectra of acetylated saponins may be caused by differences in the relative fragmentation of the M·Na^+ and M·H^+ ions between free and acetylated compounds. As a result, the acetate derivative is essential for sugar sequence information, while the free glycoside may be preferable for simple molecular weight determination. Inspection of the spectrum of non-acetylated compound (**4**) illustrated these points. Other than the quasimolecular ion, the only distinctive ions were those at m/z 413 and 421 of uncertain origin. When the acetate was used, the intensity of the M·Na^+ ion may profitably be enhanced by the external addition of sodium ion (*via* a soluble sodium salt or by dissolution of the sample in alcohol).

Besides the major acylium ions, ions apparently corresponding to the expulsion of fully acetylated internal sugars are sometimes observed. In all cases, these ions can be identified by their low intensity compared with that of the acylium ions derived from terminal sugars. A possibility remains that they represent minor structural isomers present in the sample or resulting from rearrangement during the mass spectral process.

Table 2. ^1H N.m.r. chemical shifts and average coupling constants for acetylated sugars

	Structure						
Hydrogen	(1)	(2)	(3)	(4)	(5)	(6)	(7)
2-H	4.23		4.24		4.27	4.23	4.23
3-H	4.03	3.53	4.1	3.431	4.08	4.05	4.104
12-H	5.35	5.32	5.33	5.295	5.33	5.32	5.313
18-H	2.81	2.8	2.83	2.788	2.82	2.82	2.819
23-H		4.03, 4.15		3.832, 4.271			
Me(C-24)	1.33	0.77	1.47	0.77	1.4	1.4	1.46
Me(C-25)	1.23	0.93	1.12	0.92	1.28	1.29	1.28
Me(C-26)	0.73	0.72	0.75	0.7	0.7	0.76	0.77
Me(C-27)	1.1	1.1	1.26	1.09	1.23	1.13	1.12
Me(C-29)	0.88	0.87	0.9	0.88	0.94	0.92	0.91
Me(C-30)	0.9	0.88	0.92	0.89	0.95	0.92	0.92
	Glucose	Glucose	Glucose	Glucose	Glucose		
1-H	5.59	5.57	5.59	5.54	5.56		
2-H	5.19	5.18	5.18	5.154	5.2		
3-H	5.25	5.23	5.24	5.215	5.25		
4-H	5.14	5.13	5.12	5.1	5.14		
5-H	3.78	3.8	3.79	3.762	3.78		
6-H	4.28, 4.13	4.27, 4.03	4.29, 4.05	4.245, 4.021	4.28, 4.05		
	Glucose	Galactose	Galactose	Glucose	Glucose	Glucose	Glucose
1-H	4.59	4.7	4.57	4.613	4.61	4.58	4.573
2-H	5.02	4.92	5.08	3.655	3.78	5.0	5.031
3-H	5.21	5.13	5.11	5.07	5.16	5.19	5.07
4-H	5.04	5.12	5.15	4.926	4.99	5.03	4.958
5-H	3.7	3.73	3.67	3.673	3.65	3.72	3.64
6-H	4.19, 4.04	4.33, 4.06	4.32, 4.18	4.273, 4.011	4.33, 4.18	4.19, 4.11	4.303, 4.174

Table 2 (*continued*)

Hydrogen	Structure (1)	(2)	(3)	(4)	(5)	(6)	(7)
		Arabinose		Arabinose		Arabinose	Arabinose
1-H		4.38		4.316		5.68	5.725
2-H		3.88		3.888		3.92	3.887
3-H		4.95		4.811		5.1	5.163
4-H		5.13		5.222		5.25	5.235
5-H		3.97, 3.6		3.93, 3.525		3.92, 3.72	3.918, 3.711
			Glucose		Glucose		Glucose
1-H			4.47		4.55		4.482
2-H			3.81		3.93		3.791
3-H			5.16		5.1		5.163
4-H			4.92		4.92		4.922
5-H			3.65		3.7		3.67
6-H			4.24, 4.05		4.23, 4.05		4.237, 4.026
				Arabinose		Xylose	Xylose
1-H				4.544		4.62	4.666
2-H				5.033		4.87	4.866
3-H				4.911		5.1	5.126
4-H				5.125		4.95	4.969
5-H				3.915, 3.566		4.05, 3.37	4.112, 3.35
					Rhamnose	Rhamnose	Rhamnose
1-H					4.95	4.95	4.95
2-H					5.01	5.1	5.13
3-H					5.16	5.12	5.12
4-H					5.05	3.6	3.643
5-H					4.0	3.8	3.853
6-Me					1.3	1.28	1.27

Typical coupling constants in Hz for peracetylated sugars: Glucose: $J_{1,2}$ 7.7 (ether), 8.1 (ester); $J_{2,3}$ 9.3; $J_{3,4}$ 9.4; $J_{4,5}$ 9.4; $J_{5,6a}$ 2.4, $J_{5,6b}$ 4.3; $J_{6a,6b}$ 12.4. Arabinose: $J_{1,2}$ 7; $J_{2,3}$ 9.9; $J_{3,4}$ 3.4; $J_{4,5a}$ 1.3; $J_{4,5b}$ 3.7; $J_{5a,5b}$ 13.1. Xylose: $J_{1,2}$ 7; $J_{2,3}$ 9; $J_{3,4}$ 9; $J_{4,5a}$ 9; $J_{4,5b}$ 5; $J_{5a,5b}$ 12. Rhamnose: $J_{1,2}$ 1.5 Hz (not resolved); $J_{2,3}$ 3.5; $J_{3,4}$ 10; $J_{4,5}$ 10; $J_{5,6}$ 6.2. Galactose: $J_{1,2}$ 7.5; $J_{2,3}$ 9.5; $J_{3,4}$ 3.5; $J_{4,5}$ 2.5; $J_{5,6a}$ 2.5; $J_{5,6b}$ 5; $J_{6a,6b}$ 13.

Table 3. Main observed fragments in the c.p.d.m.s.

Saponin	Sugar	*m/z*	Sugar	*m/z*	Sugar	*m/z*
(1)	Glu	331				
(2)	Glu/Gal	331	Gal–Ara	547		
(3)	Glu/Gal	331	Gal–Glu	619		
(4)	Ara Glu	259 331	Ara–Glu	547	Ara–Glu–Ara	764
(5)	Rha Glu	273 331	Rha–Glu	561	Rha–Glu–Glu	850
(6)	Ara Glu	259 331	Ara–Rha	489	Ara–Rha–Xyl	705
(7)	Ara Glu	259 331	Ara–Rha Glu–Glu	484 619	Ara–Rha–Xyl	706

Determination of the Relative Positions of the Acid and Ester Functions in Medicagenic Acid Derivatives.—Final uncertainty about structures (**1**), (**3**), and (**5**)—(**7**) regards the location of the sugar chain on one of the acid functions. This problem which can be solved by mass spectrometry, n.m.r., or chemically, was studied in depth on compound (**6**). Comparison of the methods led us to propose an alternative which has subsequently been applied to compounds (**1**), (**5**), and (**7**).

The electron-impact mass spectra of oleanene triterpenoids are dominated by a retro-Diels–Alder fragmentation which produces ions characteristic of both ends of the molecule. If the C-28 carbocylic acid is free, peaks must appear at *m/z* 203 and 248; if it is derivatized, the *m/z* 248 ion is shifted to higher mass. Unfortunately, these ions are not present in the Californium PDMS data, necessitating an alternate approach to this question of acid substitution pattern. In the event, acetylated compound (**6**) was treated with the complex BH_3·THF, in conditions where free acids are reduced faster than esters. After hydrolysis of the crude reaction mixture, bayogenin (**10**) was isolated along with a larger amount of recovered medicagenic acid (**9**). It was deduced from this experiment that C-28 is esterified in compound (**6**) but the low overall yield of the sequence and its destructive character make it impractical.

In the ^{13}C n.m.r. spectra of compounds (**1**), (**3**), and (**5**)—(**7**),

the free acid carbonyls are always found deshielded with regard to the ester carbonyls. An accurate assignment has been performed on the basis of a carbon–proton correlation through their small couplings.[16] In the case of compound (**6**) a long-range coupling was found between the acid carbonyl and one of the angular methyls, necessarily C-24. Although this experiment provided a structural proof for compound (**6**), it required large quantities of material and a long accumulation time. As an alternative, it is proposed to use lanthanide induced spectral modifications to distinguish C-23 and C-28. Addition of $Eu(fod)_3$ (0.1 mol equiv.) to a solution of acetylated (**6**) in $CDCl_3$ induced severe modifications of the ^1H n.m.r. spectrum which can be explained by multisite complexation; it is worth noting however that no complexation occurred at the ester (23 or 28) carbonyl, since the deshielded anomeric proton remained unaffected. The ^{13}C n.m.r. spectrum of the same sample also showed extensive modifications which are not easily explained. However, when the spectrum is limited to quaternary carbons, the picture is much simpler. In contrast to what was observed for the protonated sites, no noticeable shift was seen, but two carbons were missing: the acid carbonyl and C-4. This fact, which is tentatively interpreted as the consequence of Fermi interactions or of modified relaxation times, is linked to the distance between the rare earth and the carbon atoms involved. The disappearance of C-4 indicates that the free acid is C-23 and therefore that the chain of sugars is bound to C-28. The experiment has been run on compounds (**1**), (**5**), and (**7**) and the same effects were observed, leading to the proposed formulae. As these conclusions rely heavily on the correct assignments of the quaternary carbons, these have been verified by the aforementioned 'long-range' C–H correlation experiment. Resonances of C-4 and C-10 are assigned on the basis of couplings with 2-H and 3-H.

Conclusions.—Progress in the chemistry and biochemistry of saponins has so far been plagued by two main difficulties: separation/purification and structural elucidation. Despite advances in both domains based on h.p.l.c.,[17] droplet counter-current chromatography, r.l.c.c.,[18] and spectrometry including c.d.,[19] there is still need for improvement of the methods. After acetylation, saponins behave as normal organic molecules, lending themselves to conventional purification and structural investigation by modern n.m.r. and m.s. techniques. The parallel development of new, milder methods to affect clean and complete deacetylation[20] now allows the recovery of pure native substances for biological assessment.[21] Structures (**1**), (**4**), and (**6**) have been proposed independently for the major saponins of an unspecified variety of russian alfalfa.[22] The identity of these structures obtained by usual degradation techniques with those drawn from our m.s. and ^1H n.m.r. investigations shows the validity of this new methodology.

Experimental

General.—^1H N.m.r. spectra were recorded at 300 MHz on Bruker AM and AC 300 at 400 MHz on Bruker AM 400 and at 500 MHz on a Bruker WM 500 instrument. ^{13}C N.m.r. were obtained at 75 MHz on Bruker AC 300 and Varian XL-300. 2D N.m.r. experiments were performed using the standard Bruker microprograms; accumulation and processing parameters are included in the Figures. M.p.s were taken on a Reichert hot-stage microscope and were not corrected. Rotations were measured on a Perkin-Elmer 141 C automatic polarimeter. Recording conditions for the Californium plasma desorption m.s. were as follows; samples of acetylated saponins (0.1—0.5 mg) were dissolved in ethanol (~0.2 ml), ethyl acetate, or mixtures of the two. One drop of this solution was applied to the metallic side of aluminized Mylar foil stretched across a Teflon and stainless steel sample holder. During evaporation, the drop was retained in the centre of the foil by means of a glass fibre loop. The PDMS apparatus has been described in detail elsewhere.[14] A 42-cm flight tube was used with a 10 kV accelerating voltage. Data acquisition times were typically 60—90 min, and ion masses were calculated relative to the H^+ and Na^+ ions. Data were collected on a Perkin-Elmer 3220 computer using programs which were developed in Prof. Macfarlane's laboratory and modified locally (L. K. Pannell).

Saponin Extraction.—Dried powdered alfalfa root (2 kg) of the lutece variety was boiled in MeOH–water (4:1; 15 l) for 4 h. After cooling and filtration, the solvent was removed under reduced pressure and the residue was suspended in MeOH (1.75 l) at 60 °C. After filtration, the saponins were precipitated from solution by means of ether (8.75 l). The precipitate was filtered off and dissolved in water (1.3 l). The solution was then dialysed against pure water for 4 days after which time the contents of the tube were freeze-dried. The solid residue was dissolved in methanol and decolourized with carbon black. After filtration, the solution was diluted with five volumes of ether. The white saponin precipitate was filtered, and dried *in vacuo* over P_2O_5 (44 g).

Acetylation and Purification of the Saponins.—Crude saponins (8.1 g) are suspended in CH_2Cl_2 (200 ml). After addition of pyridine (8 ml), Ac_2O (8 ml), and DMAP (800 mg), the mixture was stirred for 30 h during which time progressive dissolution occurred. The solution was then washed with 10% aqueous $CuSO_4$ (2 × 250 ml), then with saturated aqueous $NaHCO_3$, and then with brine. The organic layer was dried over $NaSO_4$, filtered, and evaporated to yield crude acetylated saponin mixture (12 g) as a brownish solid. Part of this mixture (8.5 g) was chromatographed on silica gel (1 kg) under a pressure of 12 bar. The column was eluted with mixtures of $CHCl_3$–MeOH (99.5:0.5; 8 l) and (99:1; 6 l); 30 ml fractions were collected and pooled according to t.l.c. Products were visualized on t.l.c. by H_2SO_4 spray followed by heat gun heating; colours ranged from blue [(**1**), (**3**)] to red [(**5**)—(**7**)], and purple [(**2**), (**4**)]. Fractions 264—283 contained compound (**2**) and non-polar impurities (361 mg); fractions 284—305 (240 mg) were mixtures of compounds (**2**) and (**4**); fractions 306—327 contained compound (**4**) and minor impurities (448 mg). Compound (**5**) was in fractions 328—343 (1 249 mg), and in fractions 344—348 (256 mg) accompanied by compound (**3**). Fractions 349—351 (172 mg) were mixtures of compounds (**3**) and (**7**). Compound (**7**) was found almost pure in fractions 352—369 (1 190 mg); compound (**1**) was the major component of fractions 381—384 (205 mg) and the minor component of fractions 385—400 (1 211 mg), which also contained compound (**6**). Compound (**6**) was present in fractions 401—410 (291 mg). Samples suitable for n.m.r. analysis were obtained by repeated chromatography and p.l.c. of the aforementioned fractions; pooling of the fractions was then done according to 300 MHz ^1H n.m.r. spectrum of the fractions, which showed a single spot on t.l.c. Yields of the seven major saponins were *ca.* 5% (**2**), 10% (**4**), 4% (**5**), 2% (**3**), 20% (**7**), 14% (**1**), and 22% (**6**). All compounds but (**1**) were solids with no sharp m.p. (softening at 200 °C and progressive decomposition); compound (**1**) was obtained as a white powder from ether with m.p. 255 °C. Optical rotations were measured in $CHCl_3$, in a 1 dm cell; these were 28° [*c* 1, (**1**)], 20° [*c* 3.3, (**2**)], 14° [*c* 0.5, (**3**)], 31° [*c* 1.05, (**4**)], 22° [*c* 0.9, (**5**)], 0 ± 3° [*c* 1, (**6**)], and −8° [*c* 0.4, (**7**)]. P.d.m.s. quasi-molecular ions and composition: 1 186.3 [(**1**) $C_{58}H_{82}NaO_{24}$. Calc. *M*, 1 186.28]; 1 415.4 [(**2**) $C_{69}H_{98}NaO_{29}$. Calc. *M*, 1 414.52]; 1 474.5 [(**3**) $C_{70}H_{98}NaO_{32}$. Calc. *M*, 1 474.53]; 1 632.0 [(**4**) $C_{78}H_{110}NaO_{35}$. Calc. *M*, 1 630.72]; 1 705.3 [(**5**) $C_{80}H_{112}NaO_{38}$. Calc. *M*, 1 704.75]; 1 560.0 [(**6**) $C_{74}H_{104}$-

NaO_{32}. Calc. *M*, 1 560.62]; 1 850.1 [(**7**) $C_{86}H_{120}NaO_{42}$. Calc. *M*, 1 848.88].

Acknowledgements

CNRS, INRA, MRT, and the Region Champagne-Ardenne are thanked for their generous support of this work. Stimulating discussions with Dr. J. D. Connolly were greatly appreciated.

References

1 R. Tschesche and G. Wulff, *Fortschr. Chem. Org. Naturst.*, 1973, **30**, 461.
2 H. Budzikiewicz, J. M. Wilson, and C. Djerassi, *J. Am. Chem. Soc.*, 1963, **85**, 3688.
3 H. R. Schulten, T. Komori, and T. Kawasaki, *Tetrahedron*, 1977, **33**, 2595.
4 H. R. Schulten, T. Komori, T. Kawasaki, T. Okuyama, and S. Shibata, *Planta Med.*, 1982, **46**, 67.
5 K. Tori, S. Seo, A. Shimaoka, and Y. Tomita, *Tetrahedron Lett.*, 1976, 4167.
6 S. Seo, Y. Tomita, K. Tori, and Y. Yoshimura, *J. Am. Chem. Soc.*, 1980, **102**, 3331.
7 H. Ishii, K. Tori, T. Tozyo, and Y. Yoshimura, *J. Chem. Soc., Perkin Trans. 1*, 1984, 661.
8 U. Piantini, O. W. Sørensen, and R. R. Ernst, *J. Am. Chem. Soc.*, 1982, **104**, 6800.
9 D. Marion and K. Wüthrich, *Biochem. Biophys. Res. Commun.*, 1983, **113**, 967.
10 G. Wagner, *J. Magn. Reson.*, 1983, **55**, 151.
11 G. Massiot, C. Lavaud, D. Guillaume, L. Le Men-Olivier, and G. Van Binst, *J. Chem. Soc., Chem. Commun.*, 1986, 1485.
12 C. Nishino, S. Manabe, N. Enoki, T. Nagata, T. Tsushida, and E. Hamaya, *J. Chem. Soc., Chem. Commun.*, 1986, 720.
13 J. Howe and M. Jarman, *Fortschr. Chem. Org. Naturst.*, 1985, **50**, 107.
14 B. Sundqvist and R. D. Macfarlane, *Mass. Spectrom. Rev.*, 1985, **4**, 421.
15 Y. M. Yang, H. A. Lloyd, L. K. Pannell, H. N. Fales, R. D. Macfarlane, C. J. MacNeal, and Y. Ito, *Biomed. Environ. Mass. Spectro.*, 1986, **13**, 439.
16 H. Kessler, C. Griesinger, J. Zarbock, and H. R. Loosli, *J. Magn. Reson.*, 1984, **57**, 331.
17 B. Domon, A. C. Dorsaz, and K. Hostettmann, *J. Chromatogr.*, 1984, **315**, 441.
18 K. Hostettmmann, M. Hostettmann, and A. Marston, *Nat. Prod. Rep.*, 1984, 471.
19 H. Liu and K. Nakanishi, *J. Am. Chem. Soc.*, 1981, **103**, 559.
20 J. Herzig, A. Nudelman, H. E. Gottlieb, and B. Fischer, *J. Org. Chem.*, 1986, **51**, 727; N. Kunesch, C. Miet, and J. Poisson, *Tetrahedron Lett.*, 1987, **28**, 3569.
21 M. Nonaka, *Phytochemistry*, 1986, **25**, 73.
22 A. M. Timbekova and N. K. Abubakirov, *Khim. Prir. Soedin.*, 1984, 451; 1986, 607, 614.

Received 26*th October* 1987; *Paper* 7/1899