Artefact formation during acid hydrolysis of saponins from *Medicago* spp.

Aldo Tava[a*], Elisa Biazzi[a], Mariella Mella[b], Paolo Quadrelli[b] and Pinarosa Avato[c]

[a]*Consiglio per la Ricerca in Agricoltura e l'Analisi dell'Economia Agraria - Centro di Ricerca per le Produzioni Foraggere e Lattiero Casearie, CREA-FLC, v.le Piacenza 29, 26900 Lodi, Italy.*

[b]*Dipartimento di Chimica, Università degli Studi di Pavia, v.le Taramelli 12, 27100 Pavia, Italy.*

[c]*Dipartimento di Farmacia-Scienze del Farmaco, Università degli Studi di Bari Aldo Moro, via Orabona 4, 70125 Bari, Italy.*

Running Title: Artefacts from acid hydrolysis of saponins

*Corresponding author. Tel: +39 0371 404740, Fax: +39 0371 31853, E-mail: aldo.tava@crea.gov.it

ABSTRACT

Artefact compounds obtained during acid hydrolysis of saponins from *Medicago* spp. (Fabaceae), have been monitored and evaluated by GC-FID. Their identification has been performed by GC-MS and ^{1}H and ^{13}C NMR. Saponins with different substituents on the triterpenic pentacyclic aglycones were considered, and their hydrolysis products were detected and quantified during 10 hours of time course reaction. From soyasapogenol B glycoside the well known soyasapogenols B, C, D and F were obtained together with a previously undescribed sapogenol artefact identified as 3*β*,22*β*,24-trihydroxyolean-18(19)-en and named soyasapogenol H. From a zanhic acid saponin two major artefact compounds identified as 2*β*,3*β*,16*α*-trihydroxyolean-13(18)-en-23,28-dioic acid and 2*β*,3*β*,16*α*-trihydroxyolean-28,13*β*-olide-23-oic acid were obtained, together with some zanhic acid. Other compounds, detected in very small amount in the reaction mixture, were also tentatively identified based on their GC-MS and UV spectra. The other most characteristic saponins in *Medicago* spp., hederagenin, bayogenin and medicagenic acid glycosides, under acidic condition of hydrolysis, released instead the correspondent aglycones and generated a negligible amount of artefacts. Nature of artefacts and mechanism of their formation, involving a stable tertiary carbocation, is here proposed and discussed for the first time.

Keywords: Medicago spp., Triterpenic pentacyclic saponins, Acid hydrolysis, Artefact formation, Chemical structure, Soyasapogenol H, GC-MS, NMR

1. Introduction

The analysis of sapogenins is one of the method used to evaluate the saponin content in saponin-rich plants such as in *Medicago* spp. (Fabaceae) and other legumes. Sapogenins are released after acid hydrolyses of saponins, functionalised (methylated and acetylated or silylated) and then identified by GC-MS and quantified by GC-FID generally using an internal standard (Jurzysta and Jurzysta, 1978; Brawn et al., 1981; Rao et al., 1987; Tava et al., 1993; Tava et al., 1998; Tava et al., 1999; Pecetti et al., 2006). Although this method do not allow to obtain information on the sugar moieties of the saponins, it is a better and faster approach to quantify and discriminate among the different sapogenin (saponin) content, as an alternative to HPLC analyses (Nowacka et al., 1994), capillary electrophoresis separation (Tava et al., 2000), and LC-ESI-MS (Huhman et al., 2005; Kapusta et al., 2005) as well as MALDI techniques (Witkowska et al., 2008). Sapogenins should also be considered as part of the most complex structure of saponins when elucidating their chemical structure, and they are generally identified based on their gas-chromatographic properties, their GC-MS fragmentation patterns (Budzikiewicz et al., 1963) and by their NMR spectra, in which multiplicities of signals is lower compared to the corresponding spectra of saponins. Aglycone moieties were also considered in studies focussed on the biosynthesis of these specialised metabolites (Carelli et al., 2011; Tava et al., 2011; Fukushima et al., 2013; Moses et al., 2014; Biazzi et al., 2015).

Investigation on chemical structure of saponin/sapogenin is also of fundamental importance for evaluation of their properties, because differences in their chemical structure influence the structure/activity relationships (Avato et al., 2006). Moreover, it has been reported that during saponin/sapogenin manipulation a series of chemical modifications can occur on both aglycone and sugar portions (Massiot et al., 1996; Tava et al., 2003) especially in acidic environment. In particular, during hydrolysis, some saponins can produce by-product sapogenins that originated from the rearrangement of the triterpenic nucleus in presence of a strong acidic solution. It is known for example that saponins of soyasapogenol B (e.g. soyasaponin I) gave several soyasapogenol artefacts named soyasapogenols C, D and F (Jurzysta, 1984; Ireland and Dziedzic, 1986; Price et al., 1986; Mahato, 1991). These compounds have been monitored during hydrolysis and identified (in some case only partially) by means of spectroscopic techniques.

A different behaviour was instead observed for medicagenic acid, hederagenin and bayogenin glycosydes, the other most representative groups of saponins in several species of the *Medicago* genus. These saponins are more stable under hydrolysis conditions and release the corresponding sapogenins that are scarcely affected by the acid environment (Jurzysta and Jurzysta, 1978; Brawn et al., 1981; Rao et al., 1987; Massiot et al. 1988; Tava et al., 1993).

By contrast, GC investigation of hydrolysis derived products from zanhic acid glycosides, the other dominant group of saponins in *Medicago*, revealed the presence of unknown compounds in addition to intact zanhic acid. It is reported that saponins with an hydroxyl group at 16α position of the triterpenic nucleus as for zanhic acid, named echinocystic and quillaic acids, produce some artefacts in acidic conditions, that are identified as $\Delta^{13(18)}$ isomers and 28,13β-olides (Kubota et al., 1969).

In the present paper five triterpene saponins structurally different for types of aglycone and substituents were purified from *Medicago* spp., including *M. arabica* (L.) Huds., *M. arborea* L. and *M. sativa* L., and used in this study to investigate their stability under acid hydrolysis, a common procedure for the determination of the aglycone moiety. Formation of sapogenins from pure soyasapogenol B and pure zanhic acid saponins was followed during 10 hours hydrolysis in acidic conditions. Formed hydrolysis products, including a previously undescribed artefact sapogenins, were fully characterized by MS and NMR techniques. Purified glycosides of hederagenin, bayogenin and medicagenic acid were also hydrolysed under the same acidic conditions for comparison purposes. The mechanism of the artefact formation, involving a stable tertiary carbocation, is here proposed and discussed.

The present study aims to improve poor and incomplete data available in the literature on saponin stability under derivatization experimental procedures with the purpose to avoid erroneous structural attributions following their isolation and processing. In addition, knowledge of all the formed artefacts allows to exactly quantify saponins/sapogenins in saponin-rich plants by using a simple and fast GC technique. Finally, the study of the reaction mechanism of these molecule models, under specific chemical conditions might add new information to elucidate their biosynthetic intermediates.

2. Results and discussion

*2.1. Hydrolysis of soyasaponin I (**1**)*

The GC-FID and GC-MS analyses of silylated sapogenins from the acid hydrolysis of saponin **1** (Fig. 1) showed the presence of the sapogenols B (**6**), C (**11**), D (**15**) and F (**13**), together with the presence of another peak, further identified as *3β,22β,23*-trihydroxyolean-18(19)-en and named soyasapogenol H (**16**) (Fig. 2).

Although most artefacts from saponins are known since a long time, we felt however that some of the major compounds lacked adequate structural determination such as complete NMR and MS identification, and since they are still designated as unknown compounds, a reinvestigation of these genins is needed.

Compounds **6**, **11**, **13** and **15** were separated by silicagel column chromatography and analysed by NMR experiments. Compound **16** was not obtained in a pure form (see experimental) and used as a mixture of about 1:1 ratio with sapogenin F (**13**) for NMR investigations. All these data are reported in Table 1. By comparison of NMR data obtained by us with those reported in literature (although in some cases only partial data are available) (Heftmann et al., 1979, Nes et al., 1981; Kinjo et al., 1985; Baxter et al., 1990; Mahato, 1991) and interpretation of GC-MS spectra of derivatised sapogenins, all the chemical structures were attributed.

MS spectra of silylated compounds showed the molecular ion $[M]^+$ with a relative intensity of 2-8%. The trimethylsilyl derivatives of soyasapogenols B (**6a**) and C (**11a**) were easily identified by the expected retro Diels-Alder fragmentation, typical of the Δ^{12}-unsaturated oleananes (Budzikiewicz et al., 1963), with typical ions of *m/z* = 306 and *m/z* = 216, respectively. The rupture of ring C of the pentacyclic triterpenic structure was also detectable in the MS spectra of the trimethylsilyl derivatives of soyasapogenols D (**15a**) and F (**13a**), proved by the presence of ions *m/z* = 278 and 203 in both spectra, and by the rupture of ring E giving a typical fragment ions at *m/z* = 99 for **15a** and at *m/z* = 157 for **13a** that contains the substituent at C-22 ($-OCH_3$ for **15a** and $-OSi(CH_3)_3$ for **13a**).

Other relevant information on the chemical structures of these compounds came from NMR experiments. Based on the presence of seven methyl groups in all the examined sapogenins, other functional groups, such as double bonds and hydroxyl substituents on the triterpenic nucleus were evidenced in both ^1H and ^{13}C spectra.

In the ^{13}C NMR spectrum of soyasapogenol B (**6**) the two secondary alcoholic groups were registered at δ_C 77.2 and δ_C 81.5 while the primary alcoholic group at δ_C 65.6 and all confirmed also by DEPT experiments. The double bond was evidenced by the two carbon resonances at δ_C 123.9 and δ_C 145.5 in the ^{13}C NMR spectrum, while the presence of a vinyl proton triplet was revealed at δ_H 5.25 in the ^{1}H NMR spectrum (see Table 1). These signals are typical of the Δ^{12}-unsaturated oleananes (Kojima and Ogura, 1989; Baxter et al., 1990, Mahato and Kundu, 1994).

The same resonances for $\Delta^{12(13)}$ double bond were registered in both ^{1}H and ^{13}C NMR spectra of soyasapogenol C (**11**), together with signals at δ_C 135.0 and δ_C 136.8 in the ^{13}C NMR spectrum that correlate to the vinyl proton signals at δ_H 5.21 and δ_H 5.27 in the ^{1}H NMR spectrum, confirming the presence of the additional double bond $\Delta^{21(22)}$ in the molecule. Only one primary alcoholic group was found at δ_C 65.1 while the secondary alcoholic group resonated at δ_C 80.9 (Nes et al., 1981).

The carbon resonances of $\Delta^{13(18)}$ double bond of soyasapogenol F (**13**) were registered at δ_C 133.8 and δ_C 138.6, while the allyl signals, H-12 and H-19, were well evidenced at δ_H 2.71 and δ_H 1.85, and δ_H 2.34 and δ_H 1.70 in both the ^{1}H NMR and 2D NMR spectra (see Table 1). The three alcoholic carbon resonances of **13** were found at δ_C 65.5, δ_C 78.8 and δ_C 81.4.

The same $\Delta^{13(18)}$ double bond carbon resonances were registered in the NMR spectrum of soyasapogenol D (**15**) (see Table 1). The two alcoholic functions in the molecule were evidenced at δ_C 65.7 and δ_C 81.6, while the presence of the methoxy group was confirmed by the presence of the corresponding signals at δ_H 3.34 and δ_C 58.7 in the ^{1}H and ^{13}C NMR spectra, respectively.

Compound **16** was tentatively purified from the reaction mixture of sapogenins by means of different chromatographic techniques such as normal, reversed-phase and ion chromatography (see experimental), but only a mixture of about 1:1 ratio with soyasapogenol F (**13**) was obtained. This mixture was used for GC-MS and NMR experiments. Compound **16** was well separated under the GC conditions and its MS spectrum showed the same MW of soyasapogenols B (**6**) and F (**13**) ($C_{39}H_{74}Si_3O_3$, *m/z* = 674) but no ions from the retro Diels-Alder fragmentation were observed, as for soyasapogenol F (**13**). By comparison of NMR spectra of the mixture of compounds **13** and **16** with those of pure soyasapogenol F (**13**), signals of compound **16** could be well extracted. In the ^{13}C NMR spectrum the presence of 30 carbon atoms was revealed, of which seven methyl signals (δ_C 15.5, δ_C 16.8, δ_C 18.0, δ_C 18.4, δ_C 23.3, δ_C 30.4 and δ_C

32.5), two secondary alcoholic groups (δ_C 77.0 and δ_C 81.2), one primary alcoholic group (δ_C 65.3) and a double bond (δ_C 143.0 and δ_C 130.2) were evidenced in DEPT experiments. By comparison of the carbon resonances of compound **16** with those of soyasapogenols B (**6**) and F (**13**) (see Table 1), the same triterpenic pentacyclic nucleus can be deducted, but the double bond signals, registered at different resonances compared to that of soyasapogenols B (**6**) and F (**13**), let us to hypothesize its different position in the molecule. DEPT experiments clearly identified these signals as a CH and a quaternary carbon atom. In the ^{1}H NMR spectrum of the mixture of compounds **13** and **16**, a singlet vinyl proton at δ_H 4.86, not present in the ^{1}H NMR spectrum of pure soyasapogenol F (**13**) was registered, and attributed to H-19 of compound **16**. Based on these findings and by comparison of carbon resonances with data available from literature (Mahato and Kundu, 1994) the double bond in the triterpenic nucleus of **16** was attributed to $\Delta^{18(19)}$ position. This compound was identified as 3*β*,22*β*,24-trihydroxyolean-18(19)-en and named soyasapogenol H (**16**).

The mechanism of formation of soyasaponin I (**1**) artefacts is reported in Fig. 2, while the quantitative evaluation of conversion of soyasaponin I (**1**) into sapogenins during 10 hours of acid hydrolysis is reported in Fig. 3.

The total sapogenin content, espressed as μmol% (Fig. 3), increased during the first two hours of hydrolysis and then remained more or less constant for the rest of the reaction time, reaching the maximum values after 8 hours of hydrolysis, with a yield of 93.6 ± 4.5 μmol% of the total sapogenins. By contrary, a variation in the amount of the single sapogenin was instead observed during all the examined period of hydrolysis. Soyasapogenol B (**6**) showed an increase from 1 to 2 hours and then a decrease till to 10 hours hydrolysis, while a constant increase was in general observed for all the other identified artefacts (Fig. 3). These data suggest that the so formed aglycone **6** was than successively transformed into artefact sapogenols involving a formation of a stable tertiary carbocation (**12**) that, after a proton elimination, can generate sapogenins **13** and **16**. As reported by Mahato (1991), assuming all-chair conformation of **6** having D/E rings *cis*-fused, the 22*β*-axial hydroxyl experiences strong steric interactions with the 20*β*-axial methyl, in addition to that of the 17-methyl group. Migration of the double bond from the 12:13 to 13:18 position transforms the 22*β*-hydroxyl from axial to equatorial orientation, thereby lowering the 1,3-diaxial interaction. The same finding can be applied for the double bond migration from the 12:13 to 18:19 position in compound **16**. Substitution of the 22*β*-hydroxyl group in **13** with a methoxy group to

give compound **15**, may then occur as shown in Fig. 2 *via* protonation and elimination of the protonated group through participation of the 13:18 double bond which is favourably disposed to form the intermediate homoallylic carbocation **14**. A successive attack by a molecule of methanol on **14** generates a methoxy derivative **15** with retention of configuration. The presence of the so-called Winstein type homoallylic carbocation has been previously reported for other type of polycyclic compounds (Aneja et al., 1975; Cadenas et al., 2005; Ding et al., 2011). The introduction in the triterpenic nucleus of a RO- group was also confirmed by treatment of saponin **1** with CD_3OD and EtOH (see experimental) and compounds **15b** and **15c** were identified based on their molecular weight ($C_{37}H_{65}D_3Si_2O_3$ *m/z* = 619 and $C_{38}H_{70}Si_2O_3$ *m/z* = 630, respectively) and by the presence of ions at *m/z* = 102 for **15b** and *m/z* = 113 for **15c**, originated from the rupture of ring E that contains the above mentioned substituent at C-22 (-OCD_3 for **15b** and -OCH_2CH_3 for **15c**, respectively).

*2.2. Hydrolysis of zanhic acid saponin **5***

To investigate the artefacts obtained from zanhic acid saponins, compound **5** was taken as representative of zanhic acid glycosides and subjected to acid hydrolysis. The obtained aglycones (**10**, **18** and **19**) were purified by silica gel column chromatography and preparative reversed-phase HPLC. Compounds **10** (zanhic acid), **18** and **19**, were obtained in a pure grade together with several other by-products, including mono and dimethyl esters of the triterpenic dicarboxylic acid. Their structure elucidation was performed by combining NMR (Table 2), GC-MS and ESI-MS-MS data.

Compound **10**, was easily identified as zanhic acid, 2β,3β,16α-trihydroxyolean-12-en-23,28-dioic acid, on the basis of its MS fragmentation pattern, its NMR data (Table 2), and by comparison with data from the literature (Bialy et al., 1999; Kapusta et al, 2005; Tava et al., 2005).

Compound **18** showed the same molecular weight of zanhic acid, as deducted from the ESI-MS-MS and GC-MS spectra, but no ions from the retro Diels-Alder fragmentation were observed for the silyl derivative **18a**, compared to the same silyl derivative of zanhic acid **10a**. The presence of 30 carbon atoms was revealed in the ^{13}C NMR spectrum of compound **18**, of which six methyl signals at δ_C 13.0, δ_C 18.6, δ_C 19.6, δ_C 25.3, δ_C 26.7 and δ_C 33.2, three secondary alcoholic groups at δ_C 76.4, δ_C 72.1 and δ_C 71.9, and a double bond (δ_C 137.5 and δ_C 128.1) were evidenced in DEPT

experiments. Two carboxylic group signals at δ_C 181.9 and at δ_C 180.2, were also registered. By comparison of the carbon resonances of compound **18** with those of zanhic acid **10** (Table 2), the same triterpenic pentacyclic nucleus could be inferred. Only the double bond signals, upload shifted compared to the corresponding zanhic acid signals, let us to deduce its different position in the molecule. In the ^{1}H NMR spectra of **18** no vinyl protons (H-12 in zanhic acid at δ_H 5.33) were registered, and the allyl proton H-18 (at δ_H 3.04 in zanhic acid) well evidenced in the ^{1}H NMR of sapogenins possessing a 12-13 double bond, was absent. Two allyl signals, H-12 and H-19, were well evidenced at δ_H 2.71 and δ_H 1.96, and δ_H 2.43 and δ_H 1.78 in both the ^{1}H NMR and 2D NMR spectra (Table 2), indicating the 13:18 position of the double bond in the triterpenic nucleus. This compound was identified as 2β,3β,16α-trihydroxyolean-13(18)-en-23,28-dioic acid (**18**).

From ESI-MS/MS experiments compound **19** showed a molecular weight of *m/z* = 532, corresponding to a molecular formula of $C_{31}H_{48}O_7$, while from GC-MS of its derivative **19a**, a molecular ion of *m/z* = 748 was revealed corresponding to a molecular formula of $C_{40}H_{72}O_7Si_3$. As for compound **18a** no ions from retro Diels-Alder fragmentation were observed. In the ^{13}C NMR spectrum the two carbonyl resonances at δ_C 180.2 and δ_C 180.7 were registered together with three secondary alcoholic groups at δ_C 76.3, δ_C 71.9 and δ_C 70.2, that correlate with the signals at δ_H 3.96, δ_H 4.11 and δ_H 3.65 respectively, in the ^{1}H NMR spectrum. No vinyl signals were evidenced in both spectra, while six methyl (Table 2) and one methoxy resonances (δ_C 52.8) were registered and confirmed in the DEPT experiment. Moreover, a signal of a totally substituted oxygenated carbon (δ_C 93.0) was evidenced, this last being characteristic of a lactone group between C-28 and C-13 (Marx Young, et al., 1997; Martinez et al., 2015). This compound was then identified as 2*β*,3*β*,16*α*-trihydroxyolean-28,13*β*-olide-23-oic acid methyl ester (**19**).

A plausible mechanism of formation of the above artefacts from saponin **5** is reported in Fig. 4, while the quantitative evaluation of its conversion into sapogenins during the 10 hours of acid hydrolysis is reported in Fig. 5.

The total sapogenin content, espressed as μmol% (Fig. 5), slowly increased during the hydrolysis time reaching a maximum at 8-10 hours. A variation of the single sapogenin amount was also observed during all the examined period of hydrolysis, with zanhic acid (**10**) showing an increase from 1 to 2 hours and then a decrease till to 10 hours hydrolysis. A constant increase in concentration was observed for compounds **18**

and **19** (Fig. 5). These data revealed that, as for soyasapogenol B (**6**), the so formed aglycone (**10**) was successively transformed into artefact sapogenols involving a stable tertiary carbocation (**17**). By contrast, for zanhic acid no methoxy derivative was detected as for soyasapogenol B, and the formation of the isomerized compound **18** and the lactone **19** clearly indicates the inability of the 16α-hydroxyl group to act as good leaving group, presumably because of the lower stability of the corresponding secondary carbocation adjacent to the carboxylic group at C28 position. Moreover, the different geometrical features can result inapt to be stabilized through homoallylic interactions.

However, other minor byproducts (**21** and **22**) were formed in the reaction mixture after 30 hours of hydrolysis, originated from the elimination of the 16-hydroxyl group of zanhic acid (**10**). These compounds, detected in very low amount (less than 1% of the total mixture) were tentatively identified based on their GC-MS and UV spectra. An hypothesis of their formation is reported in Fig. 6. The protonation of the 16-hydroxyl group of zanhic acid (**10**) and elimination of the protonated group originate the intermediate carbocation **20** that, after a proton elimination, gave compound **21**, with a reaction mechanism very similar to that of formation of soyasapogenol C (**11**). Moreover, the presence of the carboxylic group likely could promote a different rearrangement that, after loss of CO_2, leads to the formation of conjugated 12:13, 17:18 diene (**22**), well identified based on its UV absorbance (see experimental). The presence of a conjugated diene was previously detected from the reaction products of acidic treatment of other 16-hydroxy pentacyclic compounds (Kubota et al., 1969).

*2.2. Hydrolysis of saponins **2**, **3** and **4***

Hydrolysis of saponins from hederagenin (**2**), bayogenin (**3**) and medicagenic acid (**4**) (Fig. 1) performed in the same acidic conditions, showed the presence of one compound clearly attributed to the corresponding aglycone (**7**, **8** and **9**, respectively) that, during the 10 hours of hydrolysis increased in yield reaching the maximum at 7-8 hours when all the saponin was completely hydrolysed (data not showed). As for the other group of saponins, the total amounts of aglycones slowly decreased after this time, likely due to decomposition (Tava et al., 1993). During the hydrolysis reaction, saponins possessing a carboxylic group in the triterpenic nucleus could undergo a transesterification reaction and the amount of formed methyl esters could be evaluated

by GC-MS after a direct silylation of the reaction mixtures (see experimental). Since discrete amounts of methyl esters are formed from saponins during their acid treatment (see Table 3), it is required to accomplish a complete methylation of sapogenin mixtures before sylanization or acetylation for GC analysis to avoid their incomplete derivatization. Other minor compounds were detected and tentatively identified based on their GC-MS spectra (Budzikiewicz et al., 1963), and the corresponding $\Delta^{13(18)}$ isomers (compounds **24**, **26** and **28**) and the 28,13β olides (compounds **25**, **27** and **29**) were found. These results are summarized in Table 3. For all the examined compounds a similar reaction mechanism involving a stable tertiary carbocation (**23**) can be presumed as outlined in Fig. 7.

3. Concluding remarks

Our study on the stability of *Medicago* saponins under common derivatization procedures showed that some molecular types, in particular soyasapogenols and zanhic acid, forms artefacts under acidic hydrolysis. Thus, availability of triterpenic saponins structurally different for types of aglycone and substituents allowed to investigate the reaction mechanisms involved in the formation of these by-products and characterize their structure based on the reactant saponin.

Results obtained indicated that artefacts production involves an intermediate tertiary carbocation, formed from the aglycone, which leads double bond transposition from 12:13 to 13:18 position in the triterpenic structure, affording these by-products in different amount depending on the chemical nature of the sapogenin under reaction.

The presence of an -OH group in 16 or 22 position (γ position to the double bond) of the triterpenic aglycone promotes this transposition leading to a lower steric interaction with the methyl substituents. In case of soyasapogenol B other types of rearrangement can be observed as the formation of the related isomer $\Delta^{18(19)}$ (soyasapogenol H) or the introduction in the molecule of nucleophiles (soyasapogenol D). In addition, the presence of the carboxylic group adjacent to the carbocation, as for zanhic acid, induce instead the formation of the corresponding γ-lactone.

Knowledge acquired through our study on formation of saponin artefacts under acidic hydrolysis and structural characterization of by-products formed should then be regarded of importance to discriminate among the presence of artefact compounds in saponin containing plants and avoid mistaken structural determinations.

In addition, the knowledge of nature of artefacts obtained from acid hydrolysis of a particular saponin is fundamental for the exact quantification of that saponin. The sum of all the amounts from the GC peaks gave the exact content of that particular compound from which all artefacts originates. This helps to obtain an appropriate quantitative determination of saponins/sapogenins in plant matrices, especially for the accurate determination of the most common soyasaponin I in legumes which otherwise will be underestimated.

Finally, the elucidation of the mechanism of reaction of the more unstable saponin structural types possibly allows to understand the chemical behaviour of similar natural products (eg. triterpenes with an -OH group in γ position to the double bond, such as caulophyllogenin, quillaic and echinocystic acids) which can undergo similar degradation. The characterization of these artefacts can add new evidences to understand triterpene saponins biosynthetic steps as well as to investigate the production of new molecules (including artefacts) with potential industrial interest.

4. Experimental

4.1. General experimental procedures

All pure compounds and fractions from the chromatographic steps were analysed by TLC, GC-FID, GC-MS, HPLC and NMR methods. Merck silica gel 60H were used for TLC and sapogenins were eluted with petroleum ether/$CHCl_3$/AcOH (7:2:1) or benzene/MeOH (9:1) and spots visualized by spraying with MeOH/acetic anhydride/sulfuric acid (10:1:1 v/v) followed by heating at 120°C.

GC-FID and GC-MS were performed on sapogenins as their methyl-silyl derivatives as described in Tava et al., 2005. GC-FID analyses were carried out using a Perkin-Elmer model 8500 GC equipped with a 30 m × 0.32 mm i.d., 0.25 μm, DB-5 capillary column. Injector and detector temperatures were set at 350 °C, and the oven temperature program was as follows: 90 °C for 5 min, increased at 20 °C/min to 250 °C for 1 min and then increased at 4 °C/min to 350 °C for 15 min. Samples (1 μl) were injected in the splitless mode. He was the carrier gas with a head pressure of 12.2 psi. GC-MS analyses were carried out using a Perkin-Elmer Clarus 500 GC equipped with a MS detector and a 30 m × 0.25 mm i.d., 0.25 μm, Elite-5MS capillary column using the same chromatographic conditions as for GC-FID. Mass spectra were acquired over the

50-850 amu range at 1 scan/s with an ionizing electron energy of 70 eV. Transfer line temperature was 300 °C, and the carrier gas was He at 1.2 mL/min.

Sapogenins were also submitted to HPLC analyses using a Perkin Elmer chromatograph equipped with a LC250 binary pump and DAD 235 detector. Separation was performed on a Discovery HS-C18 column (Supelco, 250 × 4.6 mm, 5μm) with the following mobile phase: solvent A: CH_3CN, 0.05% CF_3COOH; solvent B: H_2O, 1% MeOH, 0.05% CF_3COOH. Chromatographic runs were carried out under gradient elution from 50% (1 min isocratic condition) to 100% of solvent A in 20 min and remaining at 100% of A for 30 min. 20 μl of methanolic solutions (1 mg/ml) of all samples were injected. Sapogenins were eluted at 1.0 ml/min and detected by UV monitoring at 215 nm. UV spectra were collected from 190 to 350 nm.

ESI-MS-MS analyses were performed on a 1100 Series Agilent LC-MSD Trap-System VL. An Agilent Chemstation (LC-MSD trapSoftware 4.1) was used for acquisition and processing of the data. All the analyses were carried out using a ESI ion source type in the negative mode with the following settings: capillary voltage, 4000 V; nebulizer gas (N_2) 15 psi; drying gas (N_2); heated at 350°C and introduced at a flow of 5 l/min. Full scan spectra were acquired over the range of 100-2200 *m/z* with a scan time of 13000 *m/z*/sec. Automated MS-MS was performed by isolating the base peak (molecular ions) using an isolation width of 4.0 *m/z*, fragmentation amplitude of 1.0 volt, threshold set at 100 and ion charge control on, with max acquire time set at 300 ms. Samples were dissolved in MeOH:H2O (9:1) at the concentration of 20-30 ppm and injected by direct infusion at a flow rate of 10 μl/min (KDScientific Syringe Pump).

^{1}H and ^{13}C NMR were measured on a Bruker AV-300 spectrometer at the operating frequencies of 300.13 and 75.13 MHz, respectively. Sapogenins were examined as solutions in $CD_3OD/CDCl_3$ 2:1 (5-10 mg/0.5ml) in 5 mm tubes at 25°C. TMS was used as internal reference. 2D NMR experiments (H,H DQF-COSY; H,H TOCSY; H,H NOESY; H,H ROESY; H,C HSQC; H,C HMBC) were carried out on all compounds using the phase sensitive method. Based on 2D NMR analyses, assignments of ^{1}H and ^{13}C signals were obtained.

Melting points were determined using a Buchi apparatus and are uncorrect. Elemental analyses were carried out on a Carlo Erba instruments. Optical rotations were measured on a Perkin-Elmer 241 polarimeter at 25°C.

*4.2. Extraction and purification of saponins **1-5***

M. arabica (L.) Huds., *M. arborea* L. and *M. sativa* L. (Fabaceae) were grown at CREA-FLC, Lodi (45° 19'N, 9° 30' E, 81 m elevation), a location characterized by rather favourable, sub-continental climatic conditions (802 mm long-term average annual rainfall). Leaf sampling was carried out at flowering stage for all species, oven dried at 50°C, powdered and used for saponin extraction. Pure saponins **1-5** were obtained from the plant material following previously reported procedures (Bialy et al., 1999; Tava et al., 2005; Tava et al., 2009). Their purity and identity were evaluated by TLC, HPLC, NMR and ESI-MS-MS analyses (Tava et al., 2005; Tava et al., 2009). These saponins were confirmed to be: **1**: 3-*O*-*α*-L-rhamnopyranosyl(1→2)-*β*-D-galactopyranosyl(1→2)-*β*-D-glucuronopyranosyl soyasapogenol B (soyasaponin I); **2**: 3-*O*-*α*-L-arabinopyranosyl(1→2)-*β*-D-glucopyranosyl(1→2)-*α*-L-arabinopyranosyl hederagenin; **3**: 3-*O*-*α*-L-arabinopyranosyl bayogenin; **4**: 3-*O*-*β*-D-glucuronopyranosyl-28-*O*-[*α*-L-arabinopyranosyl(1→3)-*α*-L-rhamnopyranoside] medicagenic acid and **5**: 3-*O*-*β*-D-glucopyranosyl-28-*O*-[*α*-L-arabinopyranosyl(1→3)-*α*-L-rhamnopyranosyl(1→2)-*α*-L-arabinopyranoside] zanhic acid (Fig. 1).

4.3. Hydrolysis of saponins

5 mg of each pure saponin were separately treated with 30 mL of 2N HCl in MeOH:H_2O 1:1 under reflux for 8 and 30 hours, respectively. 30 ml of H_2O were than added, aglycones extracted with ethyl acetate, the organic solution treated with anhydrous Na_2SO_4 and the solvent eliminated under vacuum. The obtained aglycones were dissolved in MeOH (2 ml) and all samples except sapogenins from **1**, were divided in two subsamples. One solution was treated with CH_2N_2 and then silylated (with 0.2 ml of pyridin/hexamethyldisilazane/chlorotrimethylsilane 2:1:1 and heated at 70°C for 10 min) before GC injections. The other part of solution was directly silylated and used to evaluate the amount of the methyl esters obtained during the hydrolysis reaction. Sapogenins from **1** were directly treated with the silylation mixture. Three independent experiments were performed on each sample. Compound **1** was also separately treated with 2N HCl in CD_3OD:H_2O 1:1 and EtOH:H_2O 1:1 under reflux for 8 hours and the obtained compounds were treated as above and analysed by GC-MS.

*4.4. Artefact monitoring during acid hydrolysis of saponins **1** and **5***

Saponin **1** (10.4 mg, $C_{48}H_{78}O_{18}$, *M*r 942, 10.6 μmol) and saponin **5** (10.7 mg, $C_{52}H_{82}O_{24}$, *M*r 1090, 9.2 μmol) were separately treated with 30 mL of 2N HCl in MeOH:H_2O 1:1 under reflux. Every 2 hours, 5 ml of solution, corresponding to 1.73 mg (1.84 μmol) of saponin **1** and 1.78 mg (1.64 μmol) of saponin **5**, were sampled after cooling, 0.71 mg (1.61 μmol) of uvaol was added as internal standard and thereafter aglycones were extracted by using ethyl acetate (3 x 2 ml). The solvent was evaporated to dryness and the obtained sapogenins were methylated with CH_2N_2 and then silylated (0.2 ml of pyridin/hexamethyldisilazane/chlorotrimethylsilane 2:1:1 and heated at 70°C for 10 min) before GC injections. Hydrolysis was performed in triplicate and solutions used separately for GC-FID and GC-MS evaluation.

*4.5. Hydrolysis of saponin **1** and purification of compounds **6**, **11**, **13**, **15** and **16***

Saponin **1** (500 mg, 530.8 μmol) was treated with 300 ml of 2N HCl in MeOH:H_2O 1:1 under reflux for 8 hours. MeOH was removed under reduced pressure and aglycones were extracted by using ethyl acetate (3 × 100 ml). The organic solution was treated with anhydrous Na_2SO_4 and the solvent removed under reduced pressure to obtain 225.8 mg (493.0 μmol as soyasapogenol B equivalent, 92.9% yield) of crude sapogenin mixture. The sapogenin mixture was submitted to a 400 × 55 mm, 40-60 μm silica gel column (Merck). Fractions were eluted with $CHCl_3$ and checked by TLC developed with petroleum ether/$CHCl_3$/acetic acid (7:2:1 v/v) and toluene/MeOH (85:15 v/v), visualising the spots by spraying with MeOH/acetic anhydride/H_2SO_4 (10:1:1 v/v) followed by heating at 120°C. Fractions containing the same compounds were combined and 115.3 mg of a mixture of sapogenins **11** and **15** and 108.7 mg of a mixture of compound **6**, **13** and **16** were obtained. The first fraction was further fractionated using a silica gel column eluting with hexane/Et_2O (97:3 v/v) to obtain 8.7 mg of pure compound **11**, (soyasapogenol C) and 67.3 mg of pure compound **15** (soyasapogenol D). The second fraction was further fractionated using a silica gel column eluting with toluene/MeOH (95:5 v/v) to obtain 32.1 mg of pure compound **13**, (soyasapogenol F), 5.6 mg of pure compound **6** (soyasapogenol B) and 28.7 mg of mixture of compounds **13** and **16** in the ratio of above 1:1. Compounds **13** and **16** were tentatively separated using different chromatographic techniques such as ion

chromatography (TLC and open column chromatography using silica gel added with 20% $AgNO_3$ w/w) and reverse-phase chromatography (C18, C5 and CN stationary bonded phases HPLC columns with different solvent systems), but no separation was achieved. The purity and homogeneity of all the fractions from the chromatographic separation and the pure sapogenins were also checked by GC analyses.

*4.5.1. 3β,22β,24-trihydroxyolean-12(13)-en, soyasapogenol B (**6**)*

White solid; $C_{30}H_{50}O_3$, *M*r 458; mp 252-253°C; $[\alpha]_D^{25}$ +98.8 (*c* 0.08 MeOH); IR (KBr) ν_{max} 3420, 2950, 1630, 1075 cm^{-1}; for ^1H NMR (300 MHz, $CD_3OD/CDCl_3$ 2:1) and ^{13}C NMR (75 MHz, $CD_3OD/CDCl_3$ 2:1) see Table 1. Found: C, 78.7; H, 10.8. $C_{30}H_{50}O_3$ requires: C, 78.5; H, 11.0%.

*4.5.1.1. 3β,22β,24-trihydroxyolean-12(13)-en trimethylsilyl (**6a**)*

$C_{39}H_{74}O_3Si_3$: GC-MS *m/z* (rel. int.) 674 (2) $[M]^+$; 584 (2) $[M\text{-}90]^+$; 481 (5) $[M\text{-}90\text{-}CH_2OSi(CH_3)_3]^+$; 368 (7); 306 (92); 291 (86); 278 (18); 188 (40); 157 (70); 73 (100).

*4.5.2. 3β,24-dihydroxyolean-12(13),21(22)-dien, soyasapogenol C (**11**)*

White solid; $C_{30}H_{48}O_2$, *M*r 440; mp 240-242°C; $[\alpha]_D^{25}$ +33.8 (*c* 0.05 MeOH); IR (KBr) ν_{max} 3435, 2948, 1635, 1075 cm^{-1}; for ^1H NMR (300 MHz, $CD_3OD/CDCl_3$ 2:1) and ^{13}C NMR (75 MHz, $CD_3OD/CDCl_3$ 2:1) see Table 1. Found: C, 81.7; H, 10.9. $C_{30}H_{48}O_2$ requires: C, 81.8; H, 11.0%.

*4.5.2.1. 3β,24-dihydroxyolean-12(13),21(22)-dien trimethylsilyl (**11a**)*

$C_{36}H_{64}O_2Si_2$: GC-MS *m/z* (rel. int.) 584 (2) $[M]^+$; 494 (4) $[M\text{-}90]^+$; 391 (4) $[M\text{-}90\text{-}CH_2OSi(CH_3)_3]^+$; 368 (6); 278 (11); 216 (95); 201 (32); 188 (28); 95 (48); 73 (100).

*4.5.3. 3β,22β,24-trihydroxyolean-13(18)-en, soyasapogenol F (**13**)*

White solid; $C_{30}H_{50}O_3$, *M*r 458; mp 312-314 °C; $[\alpha]_D^{25}$ -34.8 (*c* 0.15 MeOH); IR (KBr) ν_{max} 3448, 2951, 1645, 1080 cm^{-1}; for ^1H NMR (300 MHz, $CD_3OD/CDCl_3$ 2:1) and ^{13}C NMR (75 MHz, $CD_3OD/CDCl_3$ 2:1) see Table 1. Found: C, 78.8; H, 10.8. $C_{30}H_{50}O_3$ requires: C, 78.5; H, 11.0%.

*4.5.3.1. 3β,22β,24-trihydroxyolean-13(18)-en trimethylsilyl (**13a**)*

$C_{39}H_{74}O_3Si_3$: GC-MS *m/z* (rel. int.) 674 (2) $[M]^+$; 584 (2) $[M-90]^+$; 494 (2) $[M-90-90]^+$; 481 (7) $[M-90-CH_2OSi(CH_3)_3]^+$; 391 (9) $[M-90-CH_2OSi(CH_3)_3-90]^+$; 306 (5); 278 (26); 203 (65); 157 (100); 73 (84).

*4.5.4. 3β,24-dihydroxy-22β-methoxyolean-13(18)-en, soyasapogenol D (**15**)*

White solid; $C_{31}H_{52}O_3$, *M*r 472; mp 289-291°C; $[\alpha]_D^{25}$ -51.8 (*c* 0.29 MeOH); IR (KBr) ν_{max} 3445, 2940, 1630, 1075cm^{-1}; for ^{1}H NMR (300 MHz, $CD_3OD/CDCl_3$ 2:1) and ^{13}C NMR (75 MHz, $CD_3OD/CDCl_3$ 2:1) see Table 1. Found: C, 78.6; H, 10.8. $C_{31}H_{52}O_3$ requires: C, 78.8; H, 11.1%.

*4.5.4.1. 3β,24-dihydroxy-22β-methoxyolean-13(18)-en trimethylsilyl (**15a**)*

$C_{37}H_{68}O_3Si_2$: GC-MS *m/z* (rel. int.) 616 (3) $[M]^+$; 526 (2) $[M-90]^+$; 426 (9) $[M-90-CH_2OSi(CH_3)_3]^+$; 391 (8) $[M-90-CH_2OSi(CH_3)_3-CH_3OH]^+$; 278 (29); 203 (47); 187 (46); 99 (92); 73 (100).

*4.5.4.2. 3β,24-dihydroxy-22β-deuteromethoxyolean-13(18)-en trimethylsilyl (**15b**)*

$C_{37}H_{65}D_3O_3Si_2$: GC-MS *m/z* (rel. int.) 619 (2) $[M]^+$; 530 (2) $[M-89]^+$; 426 (7) $[M-90-CH_2OSi(CH_3)_3]^+$; 391 (6) $[M-90-CH_2OSi(CH_3)_3-CD_3OH]^+$; 278 (25); 203 (28); 187 (35); 147 (45); 102 (87); 73 (100).

*4.5.4.3. 3β,24-dihydroxy-22β-ethoxyolean-13(18)-en trimethylsilyl (**15c**)*

$C_{38}H_{70}O_3Si_2$: GC-MS *m/z* (rel. int.) 630 (2) $[M]^+$; 540 (2) $[M-90]^+$; 437 (7) $[M-90-CH_2OSi(CH_3)_3]^+$; 391 (7) $[M-90-CH_2OSi(CH_3)_3-CH_3CH_2OH]^+$; 278 (23); 203 (42); 187 (31); 147 (40); 113 (92); 73 (100).

*4.5.5. 3β,22β,24-trihydroxyolean-18(19)-en, soyasapogenol H (**16**)*

For ^{1}H NMR (300 MHz, $CD_3OD/CDCl_3$ 2:1) and ^{13}C NMR (75 MHz, $CD_3OD/CDCl_3$ 2:1) see Table 1.

*4.5.5.1. 3β,22β,24-trihydroxyolean-18(19)-en trimethylsilyl (**16a**)*

$C_{39}H_{74}O_3Si_3$: GC-MS *m/z* (rel. int.) 674 (7) $[M]^+$; 585 (15) $[M-89]^+$; 481 (4) $[M-89-CH_2OSi(CH_3)_3]^+$; 391 (6) $[M-89-CH_2OSi(CH_3)_3-90]^+$; 292 (9); 277 (9); 202 (18); 187 (29); 175 (48); 147 (40); 73 (100).

*4.6. Hydrolysis of saponin **5** and purification of compounds **10**, **18** and **19***

Saponin **5** (350 mg, 321.1 μmol) was treated with 200 ml of 2N HCl in MeOH:H_2O 1:1 under reflux for 8 hours. MeOH was removed under reduced pressure and aglycones were extracted by using ethyl acetate (3 × 100 ml). The organic solution was treated with anhydrous Na_2SO_4 and the solvent removed under reduced pressure to yield 152.2 mg (293.9 μmol as zanhic acid equivalent, 91.5% yield) of crude sapogenin mixture. The sapogenin mixture was submitted to a 400 × 55 mm, 40-60 μm silica gel column (Merck). Fractions were eluted with petroleum ether/$CHCl_3$/acetic acid (7:2:0.5) and checked by TLC developed with the some solvent mixture, visualising the spots by spraying with MeOH/acetic anhydride/H_2SO_4 (10:1:1 v/v) followed by heating at 120°C. 12.1 mg of pure compound **18** were obtained. The remaining fractions from the silica gel column were further submitted to a preparative HPLC, using a Perkin Elmer liquid chromatograph equipped with a LC250 binary pump and detected by UV monitoring at 215 nm. Fractions of 200 μl (10 mg/ml of methanolic solution) were injected in a Discovery C18 column (Supelco, 10 x 250 mm, 5μm). Elution was under isocratic condition of 35% MeOH, 65% H_2O, 0.05% CF_3COOH; flow 2.0 ml/min. MeOH was removed under vacuum from the collected fractions which were then freeze-dried. In total 8.5 mg of pure zanhic acid **10** and 70.6 mg of compound **18** were obtained. Compound **19** was obtained as methyl ester (7.2 mg) after preparative HPLC of the sapogenin mixture obtained after 30 hours hydrolyses of 100 mg of saponin **5**. From a preparative HPLC a very small amount (< 1 mg) of compounds **21** and **22** were also obtained and used to confirm their structure by GC/MS.

*4.6.1. 2β,3β,16α-trihydroxyolean-12(13)-en-23,28 dioic acid, zanhic acid (**10**)*

White solid; $C_{30}H_{46}O_7$, *M*r 518; mp >320°C; $[\alpha]_D^{25}$ +65.8 (*c* 0.52 MeOH); IR (KBr) ν_{max} 3450, 2945, 1715, 1630, 1085 cm^{-1}; for ^1H NMR (300 MHz, CD_3OD) and ^{13}C NMR (75 MHz, CD_3OD) see Table 2. ESI-MS-MS *m/z* (rel. int.) 1035.2 (9) $[2M-H]^-$; 517.1 (100) $[M-H]^-$; MS^2: 499.0 (100) $[M-H-H_2O]^-$; 471.0 (12) $[M-H-H_2O-CO]^-$; 455.0 (16) $[M-H-H_2O-CO_2]^-$; 453.1 (13) $[M-H-2H_2O-CO]^-$; 437.0 (25) $[M-H-2H_2O-CO_2]^-$. Found: C, 69.8; H, 8.7%. $C_{30}H_{46}O_7$ requires: C, 69.5; H, 8.9%.

Dimethyl ester. White solid; $C_{32}H_{50}O_7$, *M*r 546; mp 261-263°C; $[\alpha]_D^{25}$ +26.0 (*c* 0.62 MeOH). Found: C, 70.1; H, 9.4%. $C_{32}H_{50}O_7$ requires: C, 70.3; H, 9.2%.

*4.6.1.1. 2β,3β,16α-trihydroxyolean-12(13)-en-23,28 dioic acid dimethyl ester trimethylsilyl (**10a**)*

$C_{41}H_{74}O_7Si_3$: GC-MS *m/z* (rel. int.) 762 (3) $[M]^+$; 703 (2) $[M-59]^+$; 672 (10) $[M-90]^+$; 613 (5) $[M-90-59]^+$; 582 (3) $[M-90-90]^+$; 523 (3) $[M-90-90-59]^+$; 411 (2); 350 (10); 260 (93); 201 (100); 187 (20); 173 (38); 147 (54); 133 (43); 73 (58).

*4.6.2. 2β,3β,16α-trihydroxyolean-13(18)-en-23,28 dioic acid (**18**)*

White solid; $C_{30}H_{46}O_7$, *M*r 518; mp >320°C; $[\alpha]_D^{25}$ +32.0 (*c* 0.51 MeOH); IR (KBr) ν_{max} 3455, 2943, 1720, 1634, 1090 cm^{-1}; for ^1H NMR (300 MHz, CD_3OD) and ^{13}C NMR (75 MHz, CD_3OD) see Table 2. ESI-MS-MS *m/z* (rel. int.) 1035.3 (13) $[2M-H]^-$; 517.1 (100) $[M-H]^-$; MS^2: 499.0 (86) $[M-H-H_2O]^-$; 473.0 (81) $[M-H-CO_2]^-$; 471.2 (13) $[M-H-H_2O-CO]^-$; 455.0 (100) $[M-H-H_2O-CO_2]^-$; 453.1 (19) $[M-H-2H_2O-CO]^-$; 437.0 (11) $[M-H-2H_2O-CO_2]^-$. Found: C, 69.2; H, 8.7. $C_{30}H_{46}O_7$ requires: C, 69.5; H, 8.9%.

Dimethyl ester. White solid; $C_{32}H_{50}O_7$, *M*r 546; mp 265-268°C; $[\alpha]_D^{25}$ -5.4 (*c* 0.65 MeOH). Found: C, 70.1; H, 9.3. $C_{32}H_{50}O_7$ requires: C, 70.3; H, 9.2%.

*4.6.2.1. 2β,3β,16α-trihydroxyolean-13(18)-en-23,28 dioic acid dimethyl ester trimethylsilyl (**18a**)*

$C_{41}H_{74}O_7Si_3$: GC-MS *m/z* (rel. int.) 762 (14) $[M]^+$; 703 (2) $[M-59]^+$; 672 (9) $[M-90]^+$; 613 (12) $[M-90-59]^+$; 582 (5) $[M-90-90]^+$; 523 (11) $[M-90-90-59]^+$; 463 (5) $[M-90-90-59-60]^+$; 411 (7); 335 (20); 321 (58); 261 (48); 247 (41); 201 (39); 187 (68); 173 (67); 147 (100); 133 (79); 73 (79).

*4.6.3. 2β,3β,16α-trihydroxyolean-28,13β-olide-23-oic acid methyl ester (**19**)*

White solid; $C_{31}H_{48}O_7$, *M*r 532; mp >320°C; $[\alpha]_D^{25}$ +15.2 (*c* 0.89 MeOH); IR (KBr) ν_{max} 3450, 2946, 1770, 1105 cm^{-1}; for ^1H NMR (300 MHz, CD_3OD) and for ^{13}C NMR (75 MHz, CD_3OD) see Table 2. ESI-MS-MS *m/z* (rel. int.) 1063.4 (100) $[2M-H]^-$; 531.0 (17) $[M-H]^-$; MS^2: 499.0 (100) $[M-H_2O-CH_3]^-$; 471.0 (3) $[M-H_2O-CH_3-CO]^-$; 455.0 (2) $[M-H_2O-CH_3-CO_2]^-$; 453.1 (3) $[M-2H_2O-CH_3-CO]^-$; 441.0 (5) $[M-H_2O-CO-CO_2]^-$. Found: C, 70.3; H, 9.2. $C_{31}H_{48}O_7$ requires: C, 69.9; H, 9.1%.

*4.6.3.1. 2β,3β,16α-trihydroxyolean-28,13β-olide-23-oic acid methyl ester trimethylsilyl (**19a**)*

$C_{40}H_{72}O_7Si_3$: GC-MS *m/z* (rel. int.) 748 (4) $[M]^+$; 689 (2) $[M-59]^+$; 658 (7) $[M-90]^+$; 599 (15) $[M-90-59]^+$; 509 (6) $[M-90-59-90]^+$; 411 (10); 334 (22); 321 (35); 306 (22); 275 (32); 244 (23); 231 (30); 187 (48); 173 (59); 147 (100); 133 (60); 75 (96).

*4.6.4. 2β,3β-dihydroxyolean-12(13),15(16)-dien-23,28-dioic acid (**21**)*

$C_{30}H_{44}O_6$, *M*r 500.

*4.6.4.1. 2β,3β-dihydroxyolean-12(13),15(16)-dien-23,28-dioic acid dimethyl ester trimethylsilyl (**21a**)*

$C_{38}H_{64}O_6Si_2$: GC-MS *m/z* (rel. int.) 762 (4) $[M]^+$; 613 (2) $[M-59]^+$; 582 (2) $[M-90]^+$; 523 (7) $[M-90-59]^+$; 463 (2) $[M-90-59-60]^+$; 433 (2) $[M-90-59-90]^+$; 411 (8); 321 (22); 260 (53); 201 (100); 187 (28); 173 (32); 147 (38); 133 (57).

*4.6.5. 2β,3β-dihydroxyolean-12(13),17(18)-dien-23-oic acid (**22**)*

$C_{29}H_{44}O_4$, *M*r 456. UV (MeOH) λ_{max} (log ε) 235 (2.85), 243 (2.84), 253 (2.66).

*4.6.5.1. 2β,3β-dihydroxyolean-12(13),17(18)-dien-23-oic acid methyl ester trimethylsilyl (**22a**)*

$C_{36}H_{62}O_4Si_2$: GC-MS *m/z* (rel. int.) 614 (23) $[M]^+$; 524 (4) $[M-90]^+$; 465 (5) $[M-90-59]^+$; 321 (5); 202 (55); 190 (48); 187 (28); 173 (22); 147 (59); 133 (38).

*4.7. GC/MS data of artefact compounds from acid hydrolysis of saponins **2**, **3** and **4***

*4.7.1. 2β,23-dihydroxyolean-13(18)-en-23-oic acid methyl ester trimethylsilyl (**24a**)*

$C_{37}H_{66}O_4Si_2$: GC-MS *m/z* (rel. int.) 630 (2) $[M]^+$; 540 (3) $[M-90]^+$; 481 (1) $[M-90-59]^+$; 278 (15); 248 (10); 201 (21); 189 (51); 173 (21); 148 (100); 133 (37).

*4.7.2. 2β,23-dihydroxyolean-28,13β-olide trimethylsilyl (**25a**)*

$C_{36}H_{64}O_4Si_2$: GC-MS *m/z* (rel. int.) 616 (2) $[M]^+$; 526 (10) $[M-90]^+$; 436 (6) $[M-90-90]^+$; 235 (8); 200 (20); 187 (23); 173 (12); 147 (58); 135 (18); 73 (100).

*4.7.3. 2β,3β,23-trihydroxyolean-13(18)-en-23-oic acid methyl ester trimethylsilyl (**26a**)*

$C_{40}H_{74}O_5Si_3$: GC-MS *m/z* (rel. int.) 718 (2) $[M]^+$; 659 (1) $[M-60]^+$; 628 (5) $[M-90]^+$; 538 (6) $[M-90-90]^+$; 525 (12); 465 (7); 275 (12); 261 (10); 191 (40); 173 (20); 147 (47); 133 (37); 73 (100).

*4.7.4. 2β,3β,23-trihydroxyolean-28,13β-olide trimethylsilyl (**27a**)*

$C_{39}H_{72}O_5Si_3$: GC-MS *m/z* (rel. int.) 704 (1) $[M]^+$; 614 (9) $[M-90]^+$; 524 (6) $[M-90-90]^+$; 511 (23); 191 (28); 147 (57); 73 (100).

*4.7.5. 2β,3β-dihydroxyolean-13(18)-en-23,28-dioic acid methyl ester trimethylsilyl (**28a**)*

$C_{38}H_{66}O_6Si_3$: GC-MS *m/z* (rel. int.) 674 (2) $[M]^+$; 584 (5) $[M-90]^+$; 525 (12) $[M-90-59]^+$; 465 (7) $[M-90-59-60]^+$; 411 (3); 335 (5); 321 (19); 261 (11); 189 (45); 147 (38); 133 (34); 73 (100).

*4.7.6. 2β,3β-dihydroxyolean-23-oic acid-28,13β-olide methyl ester trimethylsilyl (**29a**)*

$C_{37}H_{64}O_6Si_2$: GC-MS *m/z* (rel. int.) 660 (2) $[M]^+$; 570 (6) $[M-90]^+$; 511 (18) $[M-90-59]^+$; 429 (10); 411 (6); 321 (19); 275 (12); 218 (27); 189 (25); 173 (28); 147 (55); 73 (100).

Acknowledgements

This research was supported by grant from 'Consiglio per la Ricerca in Agricoltura e l'Analisi dell'Economia Agraria' (CREA), Italy, and by the Italian 'Ministero dell'Istruzione, dell'Università e della Ricerca'.

References

Aneja, R., Davies, A.P., Knaggs, J.A., 1975. Formation of a 3,5-cyclocholestan-6α-yl derivative in a nucleophilic substitution reaction of cholesterol. Tetrahedron Lett. 12, 1033-1036.

Avato, P., Bucci, R., Tava, A., Vitali, C., Rosato, A., Bialy, Z., Jurzysta, M. 2006. Antimicrobial activity of saponins from *Medicago* sp.: structure-activity relationship. Phytother. Res. 20, 454-457.

Baxter, R., Price, K.R., Fenwick, G.R., 1990. Sapogenin structure: analysis of the ^{13}C- and ^{1}H-NMR spectra of soyasapogenol B. J. Nat. Prod. 53, 289-302.

Bialy, Z., Jurzysta, M., Oleszek, W., Piacente, S., Pizza, C., 1999. Saponins in alfalfa (*Medicago sativa* L.) root and their structural elucidation. J. Agric. Food Chem. 47, 3185-3192.

Biazzi, E., Carelli, M., Tava, A., Abbruscato, P., Losini, I., Avato, P., Scotti, C., Calderini, O., 2015. CYP72A67 catalyzes a key oxidative step in *Medicago truncatula* hemolytic saponin biosynthesis. Mol. Plant 8, 1493-1506.

Brawn, P.R., Lindner, N.M., Miller, J.M., Telling, G.M., 1981. A gas chromatographic method for the determination of medicagenic acid in lucerne (alfalfa) leaf protein concentrate. J. Sci. Food Agric. 32, 1157-1162.

Budzikiewicz, H., Wilson, J.M., Djerassi, C., 1963. Mass spectrometry in structural and stereochemical problems. XXXII. Pentacyclic triterpenes. J. Am. Chem. Soc. 85, 3688-3699.

Cadenas, R.A.; Gelpi, M.E.; Mosettig, J., 2005. Nucleosteroids: synthesis of purinyl-steroid derivatives under Mitsunobu reaction conditions. J. Hetrocyclic Chem. 42, 1-3.

Carelli, M., Biazzi, E., Panara, F., Tava, A., Scaramelli, L., Porceddu, A., Graham, N., Odoardi, M., Piano, E., Arcioni, S., May, S., Scotti, C., Calderini, O., 2011. *Medicago truncatula* CYP716A12 is a multifunctional oxidase involved in the biosynthesis of hemolytic saponins. Plant Cell 23, 3070-3081.

Ding, K.; Sun, Y.S.; Tian, W.S., 2011. Total synthesis of (±)-spiniferin-1 via a polyfluoroalkanosulfonyl fluoride induced homoallilic carbocation rearrangement reaction. J. Org. Chem. 76, 1495-1498.

Fukushima, E.O., Seki, H., Sawai, S., Suzuki, M., Ohyama, K., Saito, K., Muranaka, T. 2013. Combinatorial biosynthesis of legume natural and rare triterpenoids in engineered yeast. Plant Cell Physiol. 54, 740-749.

Heftmann, E., Lundin, R.E., Haddon, W.F., Peri, I., Mor, U., Bondi, A., 1979. High-pressure liquid chromatography, nuclear magnetic resonance and mass spectra of biosynthetic soyasapogenols. J. Nat. Prod. 42, 410-416.

Huhman, D.V., Berhow, M.A., Sumner, L.W., 2005. Quantification of saponins in aerial and subterranean tissues of *Medicago truncatula.* J. Agric. Food Chem. 53, 1914-1920.

Ireland, P.A., Dziedzic, S.Z., 1986. Effect of hydrolysis on sapogenin release in soya. J. Agric. Food Chem. 34, 1037-1041.

Jurzysta, M., Jurzysta, A., 1978. Gas-liquid chromatography of trimethylsilyl ethers of soya sapogenols and medicagenic acid. J. Chromatogr. 148, 517-520.

Jurzysta, M., 1984. Transformation of soyasapogenol B into soyasapogenols C, D and F under acidic conditions. *Proc. 14th Int. Symp. Nat. Prod.* p.127.

Kapusta, I., Janda, B., Stochmal, A., Oleszek, W., 2005. Determination of saponins in aerial parts of barrel medic (*Medicago truncatula*) by liquid chromatography-electrospray ionization/mass spectrometry. J. Agric. Food Chem. 53, 7654-7660.

Kinjo, J., Miyamoto, I., Miurakami, K., Kida, K., Tomimatsu, T., Yamasaky, M., Nohara, T., 1985. Oleanane sapogenols from *Puerare* radix. Chem. Pharm. Bull. 33, 1293-1296.

Kojima, H., Ogura, H., 1989. Configurational studies on hydroxyl groups at C-2, 3 and 23 or 24 of oleanane and ursane-type triterpenes by NMR spectroscopy. Phytochemistry 28, 1703-1710.

Kubota, T., Kitatani, H., Hinoh, H., 1969. Isomerisation of quillaic acid and echinocystic acid with hydrochloric acid. Tetrahedron Lett. 10, 771-774.

Mahato, S.B., 1991. Triterpenoid saponins from *Medicago hyspida*. Phytochemistry 30, 3389-3393.

Mahato, S.B., Kundu, A.P., 1994. 13C NMR spectra of pentacyclic triterpenoids – A compilation and some salient features. Phytochemistry 37, 1517-1575.

Martinez, A., Perojil, A., Rivas, F., Parra, A., Garcia-Granados, A., Fernandez-Vivas, A. 2015. Biotransformation of oleanolic and maslinic methyl esters by *Rhizomucor miehei* CECT 2749. Phytochemistry 117, 500-508.

Marx Young, M.C., Potomati, A., Chu, E.P., Haraguchi, M., Yamamoto, M., Kawano, T. 1997. ^{13}C NMR analysis of monodesmosidic saponins from *Gomphrena macrocephala*. Phytochemistry 46, 1267-1270.

Massiot, G., Lavaud, C., Guillaume, D., Le Men-Olivier, L., 1988. Reinvestigation of sapogenins and prosapogenins from alfalfa (*Medicago sativa).* J. Agric. Food Chem. 36, 902-909.

Massiot, G., Dijoux, M.G., Lavaud, C. 1996. Saponins and artifacts. In: *Saponins Used in Food and Agricolture.* G.R. Waller and K. Yamasaki Eds. Plenum Press, New York, 183-192.

Moses, T., Pollier, J., Almagro, L., Buyst, D., Van Montagu, M., Pedreño, M.A., Martins, J.C., Thevelein, J.M., Goossens, A. 2014. Combinatorial biosynthesis of

sapogenins and saponins in *Saccharomices cerevisiae* using a C-16α hydroxylase from *Buplerum falcatum*. Proc. Natl. Acad. Sci. USA, 111, 1634-1639.

Nes, W.D., Benson, M., Heftmann, E., 1981. The location of the methylol group in sapogenol C and erythrodiol and its biosynthetic significance. Phytochemistry 20, 2299-2300.

Nowacka, J., Oleszek, W., 1994. Determination of alfalfa (*Medicago sativa*) saponins by high-performance liquid chromatography. J. Agric. Food Chem. 42, 727-730.

Pecetti, L., Tava, A., Romani, M., De Benedetto, M.G., Corsi, P., 2006. Variety and environment effects on the dynamics of saponins in lucerne (*Medicago sativa* L.). Europ. J. Agron. 25, 187-192.

Price, K.R., Fenwick, G.R., Jurzysta, M., 1986. Soyasapogenols – Separation, analysis and interconversion. J. Sci. Food Agric. 37, 1027-1034.

Rao, D., Bories, G., 1987. Simple gas chromatographic method for the determination of medicagenic acid in alfalfa (*Medicago sativa*). J. Chromatogr. 410, 169-175.

Tava, A., Oleszek, W., Jurzysta, M., Berardo, N., Odoardi, M., 1993. Alfalfa saponins and sapogenins: isolation and quantification in two different cultivars. Phytochem. Anal. 4, 269-274.

Tava, A., Pecetti, L., 1998. Hemolytic activity and saponin content in lucerne (*Medicago sativa* complex) genotypes. J. Gen. Breed. 52, 33-37.

Tava, A., Odoardi, M., Oleszek, W., 1999. Seasonal changes of saponin content in five alfalfa (*Medicago sativa*) cultivars. Agric. Medit. 129, 111-116.

Tava, A., Chiari, M., Oleszek, W., 2000. Separation of alfalfa (*M. sativa* L.) saponins as their borate complexes by capillary electrophoresis. Chapter 5. In: *Saponins in Food, Feedstuffs and Medicinal Plants.* W. Oleszek and A. Marston Eds. Kluwer Academic Publishers, Dordrecht, The Netherland, 43-56.

Tava, A., Mella, Bialy, Z., Jurzysta, M., 2003. Stability of saponins in alcoholic solutions: ester formation as artifacts. J. Agric. Food Chem. 51, 1797-1800.

Tava, A., Mella, M., Avato, P., Argentieri, M.P., Bialy, Z., Jurzysta, M., 2005. Triterpenoid glycosides from the leaves of *Medicago arborea* L. J. Agric. Food Chem. 53, 9954-9965.

Tava, A., Mella, M., Avato, P., Biazzi, E., Pecetti, L., Bialy, Z., Jurzysta, M., 2009. New triterpenic saponins from the aerial parts of *Medicago arabica* (L.) Huds. J. Agric. Food Chem. 57, 2826-2835.

Tava, A., Scotti, C., Avato, P. 2011. Biosynthesis of saponins in the genus *Medicago*. Phytochem. Rev. 10, 459-469.

Witkowska, H.E., Bialy, Z., Jurzysta, M., Waller, G.R., 2008. Analysis of saponin mixtures from alfalfa (*Medicago sativa* L.) roots using mass spectrometry with MALDI techniques. Nat. Prod. Comm. 3, 1395-1410.