This is an open access article published under an ACS AuthorChoice License, which permits copying and redistribution of the article or any adaptations for non-commercial purposes.

Organic LETTERS

Letter

pubs.acs.org/OrgLett

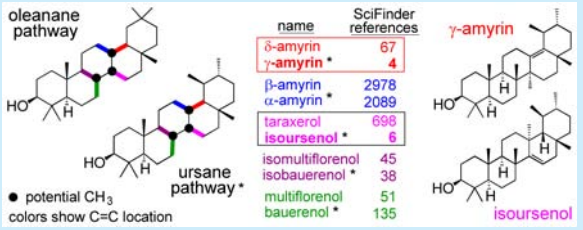
Are Isoursenol and γ-Amyrin Rare Triterpenes in Nature or Simply Overlooked by Usual Analytical Methods?

Hui Shan,[†] William K. Wilson,[†] Dorianne A. Castillo,[‡] and Seiichi P. T. Matsuda*[,†,‡]

[†]Biosciences Department and [‡]Department of Chemistry, Rice University, Houston, Texas 77005, United States

Supporting Information

ABSTRACT: Among pentacyclic triterpenes commonly found in plants, γ-amyrin and isoursenol are seldom reported and considered rare in nature. It was hypothesized that these triterpenes are instead routinely overlooked due to inadequate spectral characterization. γ-Amyrin was prepared by HCOOH isomerization of α-amyrin, and isoursenol was isolated from products of a heterologously expressed oxidosqualene cyclase. With precise NMR and GC-MS data, a metabolomics strategy was used to identify isoursenol and γ-amyrin in a wide range of plants.



Triterpenes and their triterpenoid derivatives are widely distributed in plants as specialized metabolites of medicinal, agricultural, and ecological importance.[1] More than 100 triterpene skeletons are constructed by oxidosqualene cyclases (OSCs).[2] Much of this structural diversity arises in plants from OSCs that make a mixture of oleanane (**1**−**7**) and ursane (**8**−**18**) derivatives. These pentacyclic triterpenes are generally well documented (Graphical Abstract[3]), with detailed spectral characterization (Figure 1).

Two notable exceptions are γ-amyrin[4] (urs-13(18)-en-3β-ol, **13**) and isoursenol (**15**). Prior to 1970, **13**[5] and **15**[6] were isolated and characterized by melting point and optical rotation but not by NMR or MS. Only limited spectral data for identifying **13** and **15** were reported in later references, which include a recent erroneous structure elucidation of **13** (see Figure S1).[7] Largely due to this lack of data, we[8] and others[9,10] have overlooked **15** in biological samples.

We had often anticipated the presence of **13** and **15** in metabolomics studies and as OSC products but had no NMR or GC-MS guidance to identify them in triterpene mixtures. Now serendipity has led us to **13** and **15**. Here we describe their preparation, spectral characterization, and broad distribution among higher plants.

Recently we encountered a GC-MS study of **13** in Stefan Bauer's 2002 Ph.D. dissertation.[9,11] Bauer tentatively proposed that tomato wax contains **13** on the basis of MS fragmentation and GC retention time patterns. This extensive discussion of **13** was not indexed by Chemical Abstracts, nor was **13** mentioned in Bauer's subsequent papers on tomato wax lipids.[12]

We repeated Bauer's[9] extraction and purification of tomato wax triterpenes and found a chromatographic fraction that showed a small GC-MS peak corresponding to Bauer's data for **13** (Figure 2A). Unable to locate any useful NMR data for **13**, we estimated ^{1}H and ^{13}C NMR chemical shifts from quantum mechanical calculations. Nine ^{1}H−^{13}C signal pairs from an HSQC spectrum of crude hexane extracts of tomato wax matched well with the predicted values (Table S12), which proved to be more accurate than all experimental NMR data ever reported for **13**.[13] We attempted to obtain enough material for 2D NMR characterization by repeatedly purifying the silica-gel fraction by reversed-phase HPLC. However, the near coelution of δ-amyrin (**2**) and **13** by HPLC gave at best an ∼20:1 ratio of **2**:**13** (Figure 2B). Considering Bauer's similarly futile efforts,[11] we discontinued our efforts to isolate **13** from tomato wax.

Seeking an alternative route to **13**, we envisioned synthesis from α-amyrin (**14**) via the Δ11,13(18) diene[5] or acid isomerization.[14] Given the high cost of **14**, we explored its isomerization to **13** on a low milligram scale. Treatment of **14** with triflic acid[14a] (50 mM in $CDCl_3$ at 23 °C) gave only 3% **13** in a 63:11:7:7:6:3:3 mixture of **14**, **2**, 18-epi-β-amyrin[14a] (**20**), canarenes[15] (**21**), unidentified triterpenes, **1**, and **13** (Figure S4). As in other work,[14a] our TfOH isomerizations equilibrated most of the triterpenes in Figure 3A. CF_3COOH gave at best a 50:30:10:4 mixture of **14**, canarenes, triterpene epimers, and **13**. No reaction occurred in AcOH at 90−95 °C for 20 h (Figure S5).

We then tried isomerization in formic acid. Relative to acetic acid, HCOOH is more acidic by 1 pK_a unit, has a 10-fold higher dielectric constant, and can add reversibly to olefinic bonds. In a series of HCOOH isomerizations of **14** ranging from 170 h at 40 °C to 4 h at 93 °C (Figure S6), the product consistently contained ∼4% **13** and ∼2% total **1**−**3** (from the 2% impurity of **3** in **14**). The amount of recovered **14** dropped with rising temperature from 93% to 40% due to formation of canarenes[15] (**21**) via dehydration to the C3 cation, A-ring contraction, and backbone rearrangement. The results indicate that HCOOH isomerization of **14** is limited to the ursane energy trough (magenta bars in Figure 3A), with **13** as the only isomer formed.

Based on these exploratory findings, **14** (20 mg) was isomerized in supersaturated HCOOH solution at 65 °C to a 94:6 mother-liquor mixture of **14** and **13** as formate esters **14f**

Received: June 26, 2015
Published: August 2, 2015

 © 2015 American Chemical Society

DOI: 10.1021/acs.orglett.5b01851
Org. Lett. 2015, 17, 3986−3989

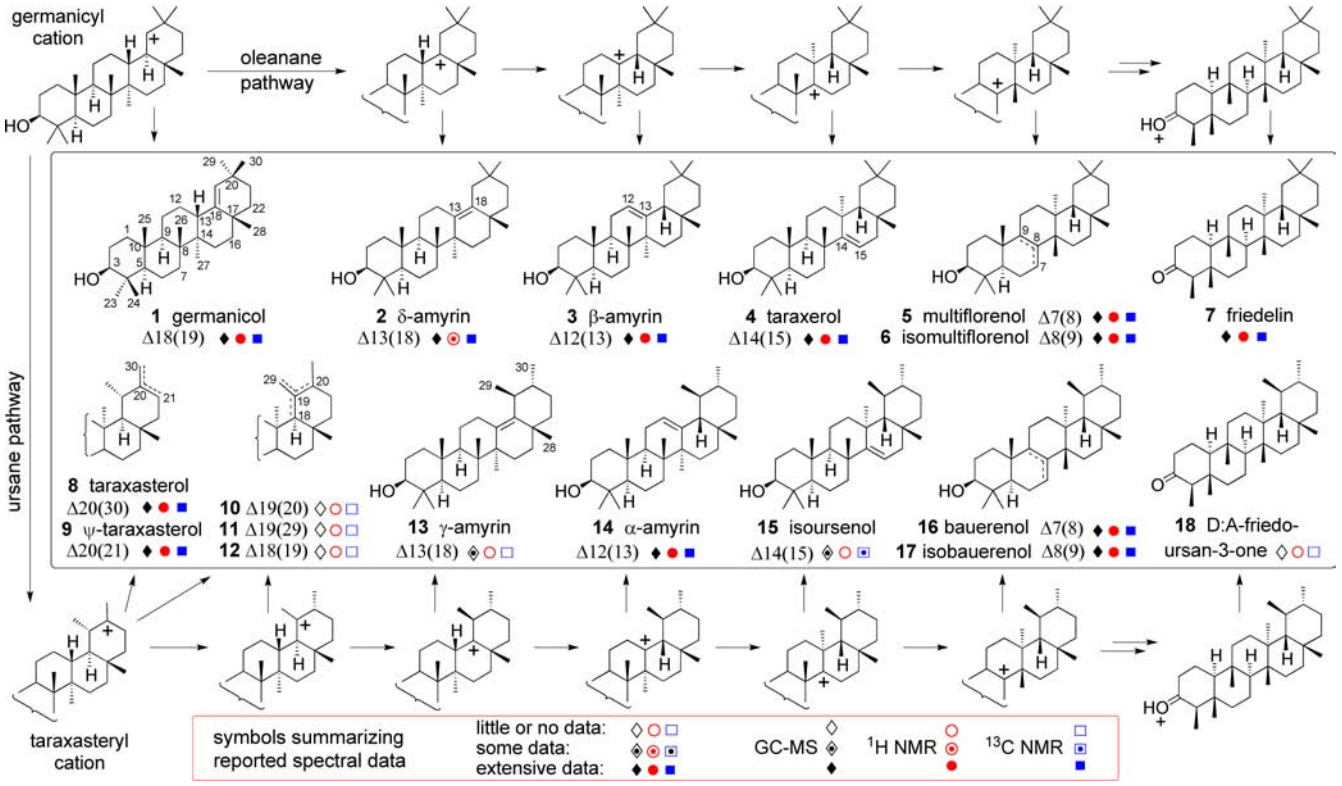


Figure 1. Mechanistic pathways to oleanane and ursane triterpenes. The extent of NMR and GC-MS characterization in the literature is summarized for each triterpene skeleton, with emphasis on NMR chemical shift precision suitable for mixture analysis (see Tables S2−S4 for details). Ursanes **10**−**12** and **18** are not well-established substances. Glutinol and other intermediates between **6** and **7** and **17** and **18** are omitted due to space limitations.

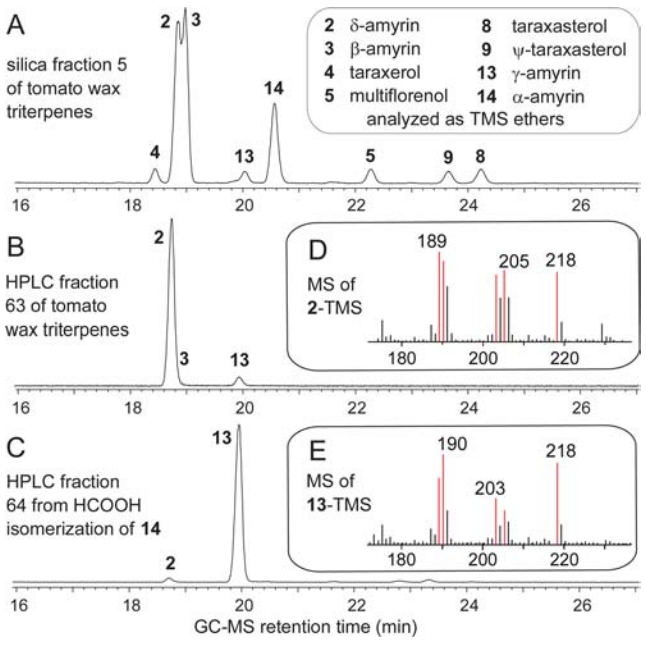


Figure 2. GC-MS total ion chromatograms (TMS ethers) from efforts to isolate **13** from tomato wax by (A) flash chromatography on silica gel and (B) C_{18} HPLC. (C) Isolation of **13** by HCOOH isomerization of **14**, followed by C_{18} HPLC. (D, E) Subtle differences in MS ion intensities[9] for *m*/*z* 189, 190, 205, and 218 distinguish pure **13** and **2** as TMS ethers but are usually obscured by overlapping impurities.

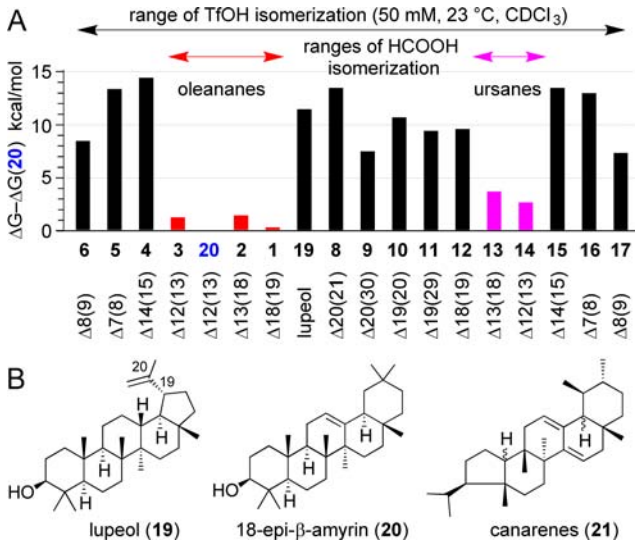


Figure 3. (A) Predicted free energies (Δ*G*, 25 °C) relative to **20** from B3PW91/6-311G(2d,p)//B3LYP/6-31G* calculations. Atom numbering is given in Figure 1. Product ratios are governed thermodynamically by energies of neutral species. Cation migration occurs readily via 1,2-shifts in TfOH but not in HCOOH. Thus, TfOH gives a wider isomerization range than HCOOH, whose largely neutral mechanism cannot bridge the ursane and oleanane energy troughs. (B) Structures of lupeol (its C20 cation being the precursor of **1**−**18**) and isomerization byproducts: 18-epi-*β*-amyrin and representative canarene isomers.

and **13f**. Partial HCOOH evaporation enriched the mother liquor to a 3:1 ratio of **14f**:**13f**, and precipitated **14f** was recycled for further isomerization. After saponification, preparative HPLC removed **14** and most other impurities, and analytical HPLC mostly removed **2** to give ∼1 mg of **13** in 93% purity (Figures 2C and S7F). NMR data at the precision required for mixture analysis[16] are given for **13** and acetate **13a** (Tables 1, S7, S8, S10,

DOI: 10.1021/acs.orglett.5b01851
Org. Lett. 2015, 17, 3986−3989

and S11). Consistency of the NMR data with calculated values (Table S12) confirmed that the olefin isomerization in HCOOH was not accompanied by epimerization.

Table 1. ^{1}H and ^{13}C NMR Data for Methyl Signals of 13 and 15[a]

	γ-amyrin (13)		isoursenol (15)	
atom	δ_C	δ_H, mult (*J*)	δ_C	δ_H, mult (*J*)
23	28.07	0.989, s	28.00	0.975, s
24	15.48	0.771, s	15.47	0.799, s
25	16.33	0.868, s	15.20	0.915, s
26	17.83	0.922, s	26.29	1.058, s
27	21.86	1.121, s	19.39	0.931, s
28	28.32	1.102, s	37.05	0.934, s
29	23.15	1.086, d (7.6)	27.48	1.021, d (6.9)
30	20.48	0.843, d (7.0)	22.46	0.977, d (6.6)

[a]Conditions: ~2 mM in $CDCl_3$, 25 °C, 800 MHz; ^{13}C chemical shifts (±0.01 ppm) were referenced to $CDCl_3$ (77.00 ppm); ^{1}H chemical shifts (±0.001 ppm) were referenced to $SiMe_4$; coarse *J* couplings are given in Hz. See Tables S7–S9 for full NMR data and fine couplings.

Like γ-amyrin, isoursenol (**15**) is an established substance. Its structure was determined in 1961–65 by synthesis from **14**,[6a,b] acidic rearrangement back to **14**,[6a] and a photooxidative reverse rearrangement.[6b] In 1966, **15** was isolated from tree leaves of an asterid and rigorously identified by isomerization to **14** and careful comparison with reported[6a] values for mp and $[\alpha]_D$. More recent references cite this early work but provide little useful spectral data.[17]

While investigating the enzymatic products of an *Arabidopsis thaliana* OSC (PEN6), we observed in a triterpene mixture a downfield olefinic NMR signal with the distinctive coupling pattern of taraxerol (dd, 8 and 3 Hz) but at a different chemical shift. Sensing that this unknown was a taraxerol analogue, we obtained quantum mechanical estimates of NMR chemical shifts for the postulated ursane derivative **15**. As with γ-amyrin, the NMR predictions confirmed the structure by their matches to the observed HSQC peaks.

After chromatographic purification, **15** and its acetate **15a** were characterized by GC-MS and NMR (Tables 1 and S9–S12), which matched known ^{13}C chemical shifts[17a] and EI-MS fragmentation[17c] for **15a**. We also reproduced the reported[6a,c] acid-catalyzed isomerization of **15** to **14** by using dilute TfOH. As in our γ-amyrin experience, a lack of useful spectral data had caused **15** to be overlooked in tomato wax,[9] PEN6 products,[10] and our own plant analyses.[8]

With precise NMR and GC-MS data now available for **13** and **15**, we looked for their presence in a variety of plants by using established methods for analyzing complex mixtures.[16a] First we revisited our existing spectra and found **13** also in PEN6 products and **15** also in tomato wax. The strongest evidence was from analysis of HSQC results (Figures 4 and S8 and Tables S6 and S15).

Among ~25 additional plants we studied,[18] **15** was found in 18 plants and **13** was in 4 of these plants (Figure 5). Our results raise the number of plants proven to make **15** from 3 to 23 and for **13** from 0 to 6. Figure 5 illustrates the broad taxonomic distribution of **15** and **13**. Most plants containing **14** also make **15** as a minor OSC byproduct. Our less frequent detection of **13** is attributable to steric hindrance of cation deprotonation (Figure S47) and to a lack of the spectral markers that allowed detection of **15** at trace levels.

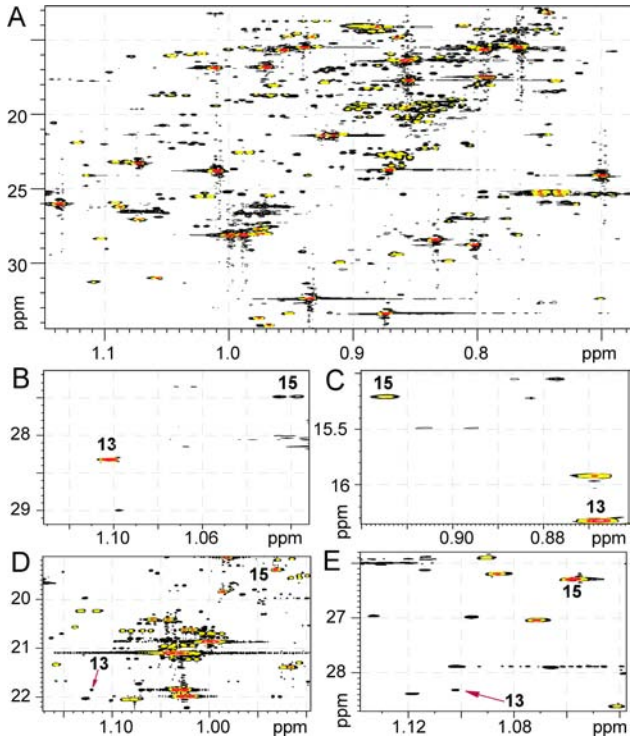


Figure 4. Detection of **13** and **15** in HSQC spectra. (A) Upfield methyl region of crude hexane extracts of tomato wax. (B and C) Expansions showing peaks for **13** and **15**. (D and E) Similar HSQC expansions show **13** and **15** in the triterpene fraction of PEN6 products.

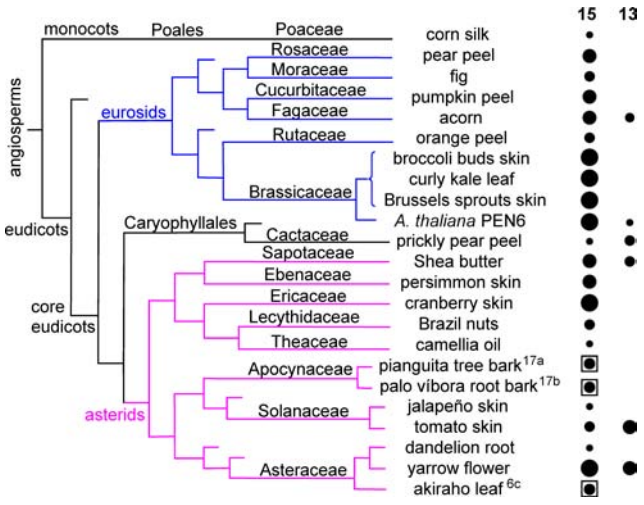


Figure 5. Observation of **15** and **13** in diverse plant species. Relative amounts of **15** and **13** are indicated by the symbol sizes; symbols in boxes denote previous isolations of **15**. In this simplified AGP III phylogeny, branch lengths have no significance. Figure S45 shows a more detailed phylogeny, with genus and species names.

We identified **13**, **15**, and other triterpenes in complex mixtures by comparing observed HSQC chemical shifts, GC retention times, and MS fragmentation against spectral databases. An Excel macro searched thousands of HSQC peaks against the precise data in Table 1 to give an objective appraisal of evidence for **13** and **15**.

Midpathway OSC byproducts **4**–**6** and **15**–**17** normally represent <1% of total triterpenes. However, kale contained nearly 1 mg/kg of **15** (3:2:1 ratio of **14**, **3**, and **15**), suggesting a

DOI: 10.1021/acs.orglett.5b01851
Org. Lett. 2015, 17, 3986–3989

possible defense function in cruciferous vegetables. We detected **13** and **15** in many other common plant foods. Some traditional diets might contain up to 1 mg/day of **13** and **15** among an abundance of biologically active triterpenoids having medicinal and health benefits.

Regarding the title question, we have shown that **13** and **15** are widespread in nature but routinely overlooked by usual analytical methods. How many more secondary metabolites are similarly overlooked in plants and microorganisms? A metabolomics approach can reveal a vast complexity of hydrophobic substances. Plant metabolic profiling, still in its infancy, might benefit from our GC-MS and HSQC strategy for mixture analysis,[16a] which can definitively identify and quantify numerous nonpolar metabolites at 1 μg levels.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.orglett.5b01851.

Experimental procedures, plant analyses, molecular modeling, GC-MS and NMR figures (PDF)

Excel macro to analyze HSQC data for the presence of specific compounds (ZIP)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: matsuda@rice.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The Robert A. Welch Foundation (C-1323) supported this work. We thank Kaylah Dunbar and Matias I. Kinzurik (Rice University) for carrying out early experiments. NMR instrumentation was supported by the W. M. Keck Foundation (University of Houston) and the John S. Dunn Sr. Gulf Coast Consortium for Magnetic Resonance (Rice University). Quantum mechanical calculations were supported by the NSF under Grants OCI-0959097 and EIA-0216467.

■ REFERENCES

(1) For reviews and leading references, see: Hill, R. A.; Connolly, J. D. *Nat. Prod. Rep.* **2015**, *32*, 273−327.

(2) (a) Xu, R.; Fazio, G. C.; Matsuda, S. P. T. *Phytochemistry* **2004**, *65*, 261−291. (b) Thimmappa, R.; Geisler, K.; Louveau, T.; O'Maille, P.; Osbourn, A. *Annu. Rev. Plant Biol.* **2014**, *65*, 225−257. (c) Phillips, D. R.; Rasbery, J. M.; Bartel, B.; Matsuda, S. P. T. *Curr. Opin. Plant Biol.* **2006**, *9*, 305−314.

(3) The Graphical Abstract shows unique SciFinder hits for the 3β-OH plus 3β-OAc triterpenes since 1970. See Table S1 for details.

(4) This trivial name fits well with the structural pattern illustrated in Figure 1: α- and β-amyrin are the Δ12 ursane and oleanane isomers, respectively, and γ- and δ-amyrin are the analogous Δ13(18) olefins. The only well-defined usage of "γ-amyrin" in the literature refers to urs-13(18)-en-3β-ol: (a) Su, Y.; Tachibana, S.; Sumimoto, M. *Mokuzai Gakkaishi* **1986**, *32*, 190−202. Other references to "γ-amyrin" are ambiguous structurally: (b) Dieterle, H.; Salomon, A.; Herzberg, E. *Arch. Pharm. Ber. Dtsch. Pharm. Ges.* **1931**, *269*, 78−87. (c) Nikiema, J. B.; Vanhaelen-Fastre, R.; Vanhaelen, M. *Planta Med.* **1997**, *63*, 486. (d) Reference 5 of Supporting Information.

(5) (a) Easton, J. D.; Manson, W.; Spring, F. S. *J. Chem. Soc.* **1953**, 943−944. (b) Beaton, J. M.; Spring, F. S.; Stevenson, R.; Strachan, W. S. *J. Chem. Soc.* **1955**, 2610−2615. (c) Yagishita, K. *Bull. Agric. Chem. Soc. Jpn.* **1959**, *23*, 217−222.

(6) (a) Laird, W.; Spring, F. S.; Stevenson, R. *J. Chem. Soc.* **1961**, 2638−2642. (b) Agata, I.; Corey, E. J.; Hortmann, A. G.; Klein, J.; Proskow, S.; Ursprung, J. J. *J. Org. Chem.* **1965**, *30*, 1698−1710. (c) Chivers, H.; Corbett, R. E.; Mitchell, R. E. M. *J. Chem. Soc. C* **1966**, 1814−1816.

(7) Hua, J.; Luo, S.-h.; Li, S.-h.; Hua, Y. *Tianran Chanwu Yanjiu Yu Kaifa* **2012**, *24*, 761−763, 831.

(8) Shan, H.; Wilson, W. K.; Phillips, D. R.; Bartel, B.; Matsuda, S. P. T. *Org. Lett.* **2008**, *10*, 1897−1900.

(9) Bauer, S. Ph.D. dissertation, Westfälischen Wilhelms-Universität: Münster, 2002.

(10) Shibuya, M.; Xiang, T.; Katsube, Y.; Otsuka, M.; Zhang, H.; Ebizuka, Y. *J. Am. Chem. Soc.* **2007**, *129*, 1450−1455.

(11) Our past familiarity with Bauer's work involved targeted searches for specific GC-MS data. The description of **13** was found during an unhurried exploration of the entire dissertation. Bauer's lack of an authentic standard or definitive literature data precluded unambiguous identification of **13**. His evidence for **13** included parallels between oleanane and ursane GC retention times. However, the retention patterns did not extend well to HPLC, and this finding raised misleading doubts about the identification of **13**. His diligent efforts could not separate **13** from **2** by HPLC (MeOH or MeCN in water) or $AgNO_3$-TLC, and separation on Ag^+-HPLC was slight.

(12) (a) Bauer, S.; Schulte, E.; Thier, H.-P. *Eur. Food Res. Technol.* **2004**, *219*, 487−491. (b) Bauer, S.; Schulte, E.; Thier, H.-P. *Eur. Food Res. Technol.* **2005**, *220*, 5−10.

(13) (a) Misra, T. N.; Singh, R. S.; Upadhyay, J.; Srivastava, R. *J. Nat. Prod.* **1984**, *47*, 368−372. (b) Reference 7. The impressive performance of DFT calculations is consistent with the typical accuracy of predicted and reported NMR chemical shifts (each roughly ±0.5 ppm for ^{13}C): (c) Plainchont, B.; Nuzillard, J.-M. *Magn. Reson. Chem.* **2013**, *51*, 54−59.

(14) Triterpene isomerization in TfOH: (a) Surendra, K.; Corey, E. J. *J. Am. Chem. Soc.* **2009**, *131*, 13928−13929. Isomerization with BF_3-etherate or H_2SO_4: (b) Ageta, H.; Shiojima, K.; Arai, Y. *Chem. Pharm. Bull.* **1987**, *35*, 2705−2716. HCOOH isomerization is less common: Reference 4b.

(15) (a) Fayez, M. B. E.; Grigor, J.; Spring, F. S.; Stevenson, R. *J. Chem. Soc.* **1955**, 3378−3383. (b) Kamdem, R. S. T.; Wafo, P.; Yousuf, S.; Ali, Z.; Adhikari, A.; Rasheed, S.; Khan, I. A.; Ngadjui, B. T.; Fun, H.-K.; Choudhary, M. I. *Org. Lett.* **2011**, *13*, 5492−5495.

(16) NMR precision of ± 0.001 and ± 0.01 for ^1H and ^{13}C (broadly reproducible for nonpolar, nonacidic, nonbasic substances) has long been essential for our spectral analyses: (a) Castillo, D. A.; Kolesnikova, M. D.; Matsuda, S. P. T. *J. Am. Chem. Soc.* **2013**, *135*, 5885−5894. (b) Herrera, J. B.; Bartel, B.; Wilson, W. K.; Matsuda, S. P. T. *Phytochemistry* **1998**, *49*, 1905−1911. (c) Wilson, W. K.; Sumpter, R. M.; Warren, J. J.; Rogers, P. S.; Ruan, B.; Schroepfer, G. J., Jr. *J. Lipid Res.* **1996**, *37*, 1529−1555.

(17) (a) Nielsen, H. B.; Hazell, A.; Hazell, R.; Ghia, F.; Torssell, K. B. G. *Phytochemistry* **1994**, *37*, 1729−1735. Although the lack of assignments and multiplicities in the ^{13}C NMR spectrum of **15a** precludes estimating HSQC methyl data for **15** or **15a**, the distinctive olefinic ^{13}C chemical shifts were used to detect acetate **15a** in a triterpene mixture: (b) Pereira, P. S.; Franca, S. d. C.; Anderson, d. O. P. V.; Breves, C. M. d. S.; Pereira, S. I. V.; Sampaio, S. V.; Nomizo, A.; Dias, D. A. *Quim. Nova* **2008**, *31*, 20−24. (c) Ito, R.; Hashimoto, I.; Masukawa, Y.; Hoshino, T. *Chem. - Eur. J.* **2013**, *19*, 17150−17158. A useful MS figure of **15a** is shown, but ^1H and ^{13}C NMR data are given only for acetone-d_6, an uncommon solvent for triterpenes.

(18) Hexane extracts of fruit skins, flowers, roots, or nonsaponifiable lipids contained a mixture of phytosterols, waxes, triterpenoids, and other metabolites. Flash chromatography gave a triterpene alcohol fraction that was enriched in **13** and **15** by HPLC and analyzed by GC-MS and NMR.

DOI: 10.1021/acs.orglett.5b01851
Org. Lett. 2015, 17, 3986−3989