Structure and Origin of Two Triterpene-derived Aromatic Hydrocarbons in Messel Shale

Philippe Schaeffer, Jean-M. Trendel and Pierre Albrecht*

Institut de Chimie, Université Louis Pasteur, URA 31 du CNRS, 1 rue Blaise Pascal, 67008 Strasbourg Cedex, France

Two triterpene-derived aromatic hydrocarbons of the gammacerane-type have been isolated from Messel shale; structural (MS, NMR) and carbon isotopic data indicate that these biological markers are derived from tetrahymanol from protozoa or purple phototrophic bacteria, and from a bacterial hopanoid.

Unequivocal structural elucidation of biological markers occurring in sediments can give valuable clues to the origin of organic matter and its transformations in the subsurface.[1] In particular, pentacyclic triterpenes commonly lead to partially aromatised hydrocarbons which often appear as major constituents of the aromatic fractions.[1–3]

In the course of a detailed study of the Eocene Messel shale (Germany), GC–MS analysis of the aromatic fraction revealed the presence of three compounds **1–3**; the mass spectra† suggested they are closely related aromatic hydrocarbons, possibly a homologous series. In addition to molecular ions at *m/z* 378 (**1**, $C_{28}H_{42}$), 392 (**2**, $C_{29}H_{44}$), 406 (**3**, $C_{30}H_{46}$) and very weak ions at ($M^+ - 15$) and 191, the fragmentation patterns are essentially characterised by a triplet at *m/z* [145 + (*n* × 14)] (base peak), [157 + (*n* × 14)] and [172 + (*n* × 14)] (*n* = 0–2). Furthermore, the mass spectrum of **1** is almost identical to that of 24,25-dinorlupa-1,3,5(10)-triene **5**,[2] suggesting a common partial aromatic structure. For all these compounds, the prominent base peak is thought to result from the formation of an indanyl cation derivative.[4] These preliminary data strongly suggest monoaromatic triterpenoid hydrocarbons with a gammacerane-type basic skeleton.

This hypothesis was confirmed by exhaustive NMR investigations (Table 1) of hydrocarbons **1** and **3** after their isolation using a multi-step reversed phase HPLC fractionation.‡ For both compounds, a common basic backbone was established from the long range ($^{2,3}J_{CH}$ 7 Hz) ^1H–^{13}C correlation network (HMBC spectrum) and the carbon sequence was completed from the ^1H–^1H COSY data. For **3**, the precise location of the isopropyl group, at C(20) or C(21), remained to be determined, however, connectivities obtained by NOESY (τ_m = 1.5 s) unambiguously indicate that the isopropyl group must be attached at C(21) (Fig. 1). Stereochemical information (α or β relative orientation of methyl groups and protons) was obtained from the NOESY data and from proton coupling patterns. Furthermore, the chemical shifts of the proton and carbon resonances, virtually identical for the non-aromatic part of both products and reflecting the presence or absence of γ-*gauche* effects, firmly support a *trans-transoid-trans* stereochemistry for rings A–D. Thus, for example, the high-field ^{13}C signals (δ < 20) of the ring-junction methyl groups are in agreement with their axial orientation (presence of γ-*gauche* effects). Conversely, γ-*gauche* interactions with the axial 24-, 25- and 26-methyl groups explain the high-field ^{13}C shifts (δ < 22) of the methylene carbon atoms C(2), C(6) and C(11) and the deshielding of their axial protons (δ^1H > 1.4). Additional evidence for the structural assignment of **1** and **3** are the strong similarities observed between the ^{13}C and ^1H chemical shifts of rings A, B and partially C with those of the hopane (*A'*-neogammacerane) skeleton **6**,[5] and for **1**, the ^1H signals of the aromatic nucleus and its surroundings are almost identical to those of 24,25-dinorlupa-1,3,5(10)-triene **5**.[2]

Compound **2** could not be isolated with high enough purity for NMR studies, therefore the structure is only tentatively proposed, but is consistent with a probable common origin with **3** as shown by carbon isotopic data (see below).

It seems quite clear that the parent structure of hydrocarbon **1** must be tetrahymanol **4** (gammaceran-21α-ol), a triterpenol widespread in marine sediments, and is most likely from a protozoon source,[6] although it has also been reported to come from a purple phototrophic bacterium.[7] Tetrahymanol bears many structural similarities with 3-oxygenated triterpenes from higher plants and its geological fate could, therefore, be similar to that of higher plant triterpene alcohols for which microbially mediated aromatization processes starting in ring A and triggered by the elimination of the oxygen function are the most general transformations operating in sediments at the earlier stage of diagenesis.[2]

Since structure **3** (and possibly **2**) cannot be directly related to a known biomolecule bearing an adequate hydrocarbon framework, it can be viewed either as the representative of a new class of molecular fossils the precursors of which are yet unknown,

Table 1 ^1H and ^{13}C NMR data[a] for 28,30-dinorgammacera-17,19,21-triene **1** and 21-propyl-28,29,30-trinorgammacera-17,19,21-triene **3**

	δ_C		δ_H					δ_C		δ_H			
Carbon	**1**	**3**[b]	**1**		**3**		Carbon	**1**	**3**[b]	**1**		**3**	
1	40.83	40.7	1.74 (β)	0.83 (α)	1.73 (β)	0.84 (α)	16	25.18	27.7	2.64	2.58	2.76	2.70
2	19.14[c]	~19.0	1.65 (β)	1.42 (α)	1.64 (β)	1.40 (α)	17	135.42	136.7				
3	42.51	42.4	1.37 (β)	1.16 (α)	1.35 (β)	1.15 (α)	18	141.58	139.2				
4	33.63	33.6					19	123.56	125.5	7.10 (d, *J* 8.0 Hz)		7.13 (d, *J* 8.0 Hz)	
5	57.07	57.0	0.78		0.80		20	125.49	124.0	7.01 (dd, *J* 8.0, 8.0 Hz)		6.96 (dd, *J* 8.0, 1.5 Hz)	
6	19.00[c]	~19.0	1.56 (α)	1.40 (β)	1.55 (α)	1.41 (β)	21	126.95	135.9	6.94 (d, *J* 8.0 Hz)			
7	34.52	34.4	1.57 (α)	1.46 (β)	1.56 (α)	1.43 (β)	22	135.93	126.5			6.91 (s)	
8	41.07	41.3					23	33.50	33.5	0.866 (s)		0.860 (s)	
9	51.33	51.4	1.48		1.46		24	21.68	21.6	0.831 (s)		0.825 (s)	
10	37.95	37.9					25	16.73	16.6	0.915 (d, *J* 0.9 Hz)		0.904 (s)	
11	21.53	21.5	1.72 (α)	1.44 (β)	1.71	1.44	26	15.94	15.8	1.070 (d, *J* 0.7 Hz)		1.060 (s)	
12	26.60	26.2	2.32 (β)	1.31 (α)	2.28 (β)	1.30 (α)	27	14.39	14.8	0.787 (s)		0.781 (s)	
13	40.53	39.9	2.83		2.79		29	19.74		2.198 (s)			
14	39.82	40.4					1′		24.3			1.202 (d, *J* 6.9 Hz)	
15	29.34	29.6	1.78 (β)	1.55 (α)	1.80 (β)	1.44 (α)	2′		34.0			2.81 (sept, *J* 6.9 Hz)	
							3′		24.3			1.202 (d, *J* 6.9 Hz)	

[a] Bruker ARX 500; 500.1 MHz for ^1H, 125.8 MHz for ^{13}C; CD_2Cl_2; $SiMe_4$ standard. [b] ^{13}C chemical shifts from HMBC and HMQC spectra, ^{13}C and DEPT spectra were not recorded due to the low amounts of **3** available. [c] Assignments interchangeable.

or, more reasonably, as resulting from the rearrangement of a known parent triterpenoid. With regard to the latter hypothesis, only the hopane skeleton **6**, widespread in bacteria,[8] seems to possess structural features which could allow formation of **3**. This would require expansion of ring E favoured by a functionalised 28-methyl group located at C(18) (and in this case *via* a spiro intermediate) or C(17) [after an acid catalysed methyl shift from C(18)[9]]. Further evidence supporting a common procaryotic origin for **2** and **3** is provided by their carbon isotopic composition. In the Messel shale, isotopic differences between the triterpene-derived hydrocarbon series clearly reflect the diversity of the sources of the organic matter,[10] and the δ values§ measured for **2** (−41.5) and **3** (−42.5) fall within the range −34 to −66, specific to the bacterial hopanoids. Compound **1** (−33.1) is less depleted in ^{13}C, implying a different origin.

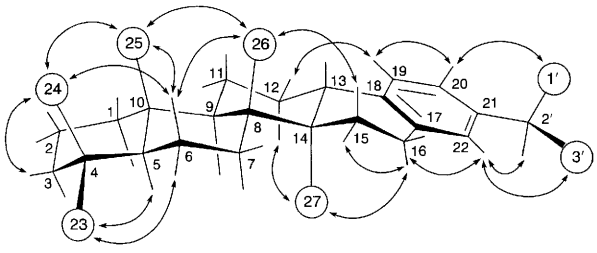
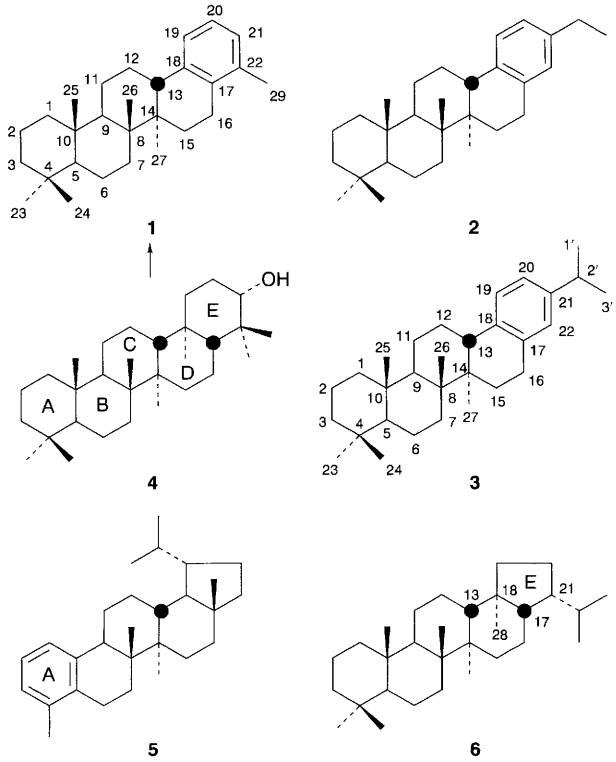


Fig. 1 3-Dimensional drawing of **3** showing the most relevant NOEs observed



We thank the Conseil Régional de la Région Alsace and the Centre National de la Recherche Scientific for financial support; R. Graff for NMR assistance; E. Mastio and P. Wehrung for MS analyses.

Received, 1st March 1995; Com. 5/01267H

Footnotes

† *MS data* for **1**: Finnigan MAT TSQ70, EI (70 eV), *m/z* (rel. int.) 378 (M^+, 9%), 363(0.5), 232(3), 197(2), 191(5), 172(19), 157(14), 145(100), 123(4), 95(6).

For **2**: 392 (M^+, 11%), 377(0.4), 211(2), 191(3), 186(24), 171(11), 159(100), 143(5), 129(4), 95(4).

For **3**: 406 (M^+, 9%), 391(0.5), 225(1), 200(23), 185(11), 173(100), 157(3), 131(4), 95(4).

‡ Extraction of the Messel shale (40 kg) and fractionation of the organic extract (35 g kg^{-1} shale) has been previously described.[3] Following preparative HPLC, the fractions containing the target products were successively subjected to semipreparative reversed phase HPLC (DuPont 250 × 9.4 mm; Zorbax ODS, 7 μm, 80 Å; methanol–chloroform 80:20 or 85:15, 5 ml min^{-1}) and to analytical reversed phase HPLC (Baker 250 × 4.6 mm; Bakerbond WP Octadecyl, 5 μm, 300 Å; methanol–chloroform 90:10 or 85:15, 1 ml min^{-1}), leading to the isolation of hydrocarbons **1** (1.5 mg) and **3** (0.2–0.3 mg) with at least 90% purity by GC.

§ Stable carbon isotope ratios were measured for single components on a Finnigan MAT 252 isotope-ratio monitoring GC–MS instrument. Carbon isotopic abundance is expressed by the $\delta^{13}C$ value = $10^3[(R_x - R_s)/R_s]$, where $R = {}^{13}C/{}^{12}C$, x designates sample, s the PeeDee Belemnite standard and R_s = 0.0112372. Mean values from at least three replicate measurements.

References

1 K. E. Peters and J. M. Moldowan, *The Biomarker Guide*, Prentice Hall, Englewood Cliffs, 1993.
2 G. A. Wolff, J. M. Trendel and P. Albrecht, *Tetrahedron*, 1989, **45**, 6721.
3 V. Hauke, R. Graff, P. Wehrung, J. M. Trendel, P. Albrecht, L. Schwark, B. J. Keely and T. M. Peakman, *Tetrahedron*, 1992, **48**, 3915.
4 C. R. Enzell, *Tetrahedron Lett.*, 1966, 2135; H. E. Audier, S. Bory, G. Defaye, M. Fétizon and G. Moreau, *Bull. Soc. Chim. Fr.*, 1966, 3181; D. A. Azevedo, F. R. Aquino Neto and B. R. T. Simoneit, *Org. Mass Spectrom.*, 1990, **25**, 475.
5 J. M. Moldowan, F. J. Fago, R. M. K. Carlson, D. C. Young, G. van Duyne, J. Clardy, M. Schoell, C. T. Pillinger and D. S. Watt, *Geochim. Cosmochim. Acta*, 1991, **55**, 3333.
6 H. R. Harvey and G. B. McManus, *Geochim. Cosmochim. Acta*, 1991, **55**, 3387, and references cited therein.
7 G. Kleemann, K. Poralla, G. Englert, H. Kjøsen, S. Liaaen-Jensen, S. Neunlist and M. Rohmer, *J. Gen. Microbiol.*, 1990, **136**, 2551.
8 G. Ourisson and M. Rohmer, *Acc. Chem. Res.*, 1992, **25**, 403.
9 V. Hauke, J. M. Trendel, P. Albrecht and J. Connan, *Tetrahedron Lett.*, 1994, **35**, 2227.
10 K. H. Freeman, J. M. Hayes, J.-M. Trendel and P. Albrecht, *Nature*, 1990, **343**, 254; K. H. Freeman, C. J. Boreham, R. E. Summons and J. M. Hayes, *Org. Geochem.*, 1994, **21**, 1037.