Identification of Long-chain Tricyclic Terpene Hydrocarbons (C_{21}—C_{30}) in Geological Samples

Denis Heissler, Ruben Ocampo, Pierre Albrecht,* Jean-Jacques Riehl, and Guy Ourisson

Laboratoire de Chimie Organique des Substances Naturelles, LA 31 associé au CNRS, Départment de Chimie, Université Louis Pasteur, 1, rue Blaise Pascal, 67008 Strasbourg, France

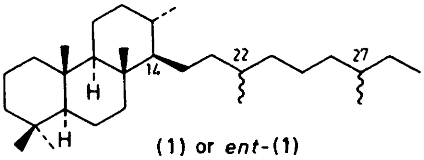
The structure of a series of long-chain tricyclic terpene hydrocarbons occurring in sediments and petroleums has been confirmed by synthesis of the C_{21}—C_{30} members; these compounds could result from a yet unknown cyclisation of polyprenols in micro-organisms.

There has been increasing interest in recent years in a novel class of tricyclic terpene hydrocarbons which occur widely in sediments and petroleums.[1] We have reported previously the identification, by synthesis, of the lower members (C_{19}, C_{20}) of this series and have thus established the relative stereochemistry of the tricyclic ring system, as well as the most likely position of the isoprenoid side chain.[2,3] Independent evidence for such a structure was also obtained from spectroscopic data in the case of the Athabasca oil sands,[4,5] where a relationship has been established between the tricyclic hydrocarbons and polar counterparts.[5,6]

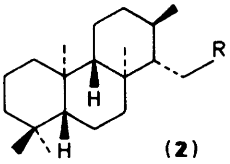
The tricyclic terpene skeleton, as yet unknown beyond C_{25} in living organisms, was found to culminate in the C_{30} homologue (**1**) in geological samples. However, evidence for a further extension of the side chain has been presented recently in the case of two petroleums.[3,7]

We now report the identification of several higher homologues (C_{21}—C_{30}) by comparison with synthetic tricyclic terpenes (**2**), thus establishing the structure of the isoprenoid side chain in these geolipids.

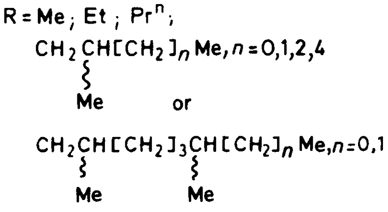
The syntheses of compounds (**2**) were performed with methyl copalate (**3**) as starting material† according to Scheme 1, except for the C_{21} homologue which was obtained earlier by a modified procedure.‡



(**1**) or *ent*-(**1**)



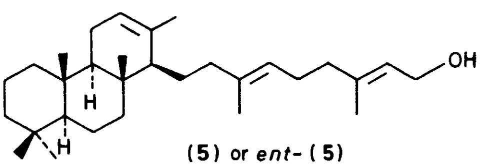
(**2**)



$R = Me;\ Et;\ Pr^n;$

$CH_2\underset{Me}{CH}[CH_2]_nMe, n = 0,1,2,4$ or

$CH_2\underset{Me}{CH}[CH_2]_3\underset{Me}{CH}[CH_2]_nMe, n = 0,1$



(**5**) or *ent*-(**5**)

---

† Copalic acid was separated from Brazilian Copaiba oil, as described by T. Nakano and C. Djerassi, *J. Org. Chem.*, 1961, **26**, 167. See also ref. 15. The absolute stereochemistry of the geoterpenes is undetermined: (**1**) or *ent*-(**1**), probably not *rac*-(**1**). (However for Lewis acid-induced formation of a tricyclic terpenoid from a polyprenol see M. Nishizawa, H. Takenaka, and Y. Hayashi, *Chem. Lett.*, 1983, 1459).

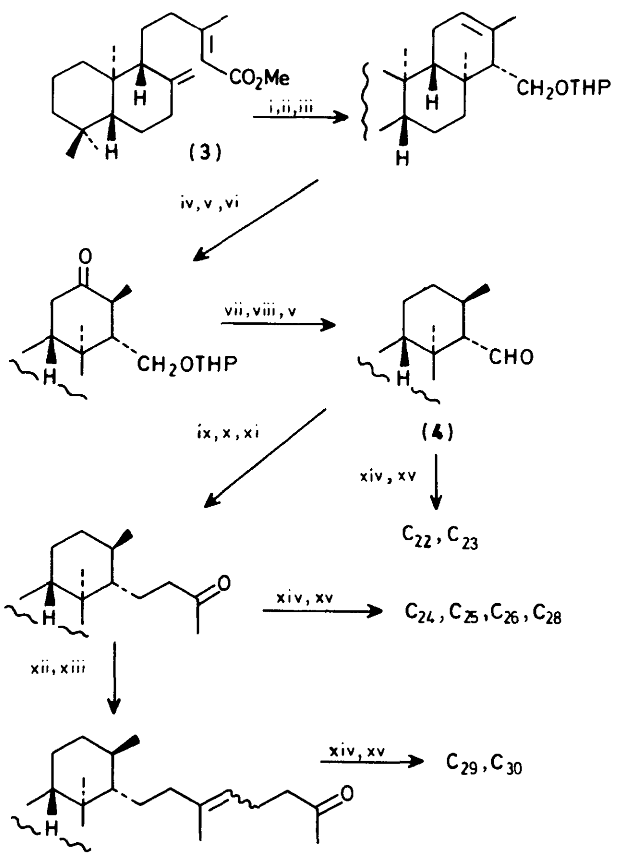
‡ The C_{21} terpene was obtained in 44% overall yield starting from the alcohol precursor to aldehyde (**4**), by the following sequence of reactions: i, $MeSO_2Cl$, Et_3N, CH_2Cl_2; ii, NaCN, DMSO; iii, di-isobutylaluminium hydride; 10% aqueous HCl; iv, $LiAlH_4$; v, $MeSO_2Cl$, Et_3N, CH_2Cl_2; vi, $LiBHEt_3$.

The tricyclic terpenes above C_{24} were obtained as a nearly 1 : 1 mixture of diastereoisomers at C-22 which could be resolved by capillary column g.c., except for the C_{25} compound which appeared as a broad peak. No further splitting was observed on our columns in the case of the C_{30} compound. Except in a few cases,[§] the C_{22} and C_{27} homologues are usually very minor components of geological samples, owing to branching at C-22 and C-27.[2,5] Their formation can be attributed to the cleavage of two carbon–carbon bonds and is therefore unfavoured, a feature commonly observed for isoprenoid alkane species in the subsurface.[8,9,¶]

The synthetic compounds were identical to their equivalent geological samples as shown by comparison of their mass spectra and cochromatography on glass capillary columns coated with two different phases (SE 30 and OV 1701; 25 m × 0.25 mm). The coinjections were performed on the alkanes from a biodegraded petroleum (Emeraude, Congo)[3] with computerised gas chromatography–mass spectrometry (LKB 9000 S GC-MS) monitoring of the m/z 191 ion (base peak) and the molecular ions. The mass spectral and n.m.r. data for the synthetic compounds were also in good agreement with those obtained previously from the Athabasca oil sands.[4,5]

Furthermore, our results confirm that the higher homologues ($>C_{24}$) of the tricyclic terpenes do indeed occur as 1 : 1 mixtures of diastereoisomers at C-22 in geological samples, as suspected previously.[2,5] The configuration of these compounds must therefore have undergone randomisation at C-22 during maturation in the subsurface, a process observed in other terpenoid series.[10]

As these long-chain tricyclic terpene hydrocarbons have not yet been recognised in living organisms, hypotheses concerning their origin can only be speculative. Because of their widespread occurrence in subsurface samples, other than in unpolluted sediments at an early stage of diagenesis, we have recently postulated that they arise from micro-organisms where their potential precursor, tricyclohexaprenol (**5**), could have been formed by the cyclisation of a regular hexaprenol.[2] Subsequent transformations in the geological environment would have led to the tricyclic alkane series. The eventual confirmation of this hypothesis would be another example of the predictive aspect of geochemical studies,[11] similar to the case of the pentacyclic hopanoids first found as ubiquitous geolipids before being discovered as widespread membrane constituents of micro-organisms.[12]



Scheme 1. Synthesis of the C_{22}—C_{30} tricyclic terpene hydrocarbons (THP = tetrahydropyran-2-yl). i, HCO_2H, room temp., 12 h, (78% yield) (ref. 14, 15); ii, $LiAlH_4$, Et_2O, 4 h (94%); iii, dihydropyran, CH_2Cl_2, camphor-10-sulphonic acid (100%); iv, BH_3–SMe_2, tetrahydrofuran, 4 h, followed by EtOH, 3 M aqueous NaOH, 30% H_2O_2, reflux, 1 h (88%); v, CrO_3–2 pyridine,CH_2Cl_2; vi, MeONa, MeOH, (97%); vii, N_2H_4, 1 H_2O, p-TsOH, (Ts = Tosyl), diethylene glycol, 130 °C, 4 h followed by KOH, 210 °C, 3 h (90%) (ref. 16); viii, 10% aqueous HCl, MeOH, hexane, (82%); ix, $(EtO)_2P(O)CHNaCN$, dimethyl sulphoxide (DMSO), (87%) (ref. 17); x, Mg, MeOH, 4 h (73%) (ref. 18); xi, MeLi, Et_2O, 0 °C, followed by 10% aqueous HCl, (67%), 19% recovered starting material; xii, $Ph_3P{=}CH[CH_2]_2C(OCH_2CH_2O)Me$, DMSO, (90%) (ref. 19); xiii, p-TsOH, EtOH, H_2O, 48 h (83%); xiv, Wittig reaction; xv, H_2, 10% Pd–C, MeOH, EtOAc.

The recent detection of higher homologues of the tricyclic series in two petroleums[3,7] (up to C_{45} in one of them[7]) indicates that a part at least of the geological terpenes must come from higher polyprenols. The latter are known as universal cell constituents. In bacteria the role of polyprenols (in particular undecaprenols, C_{55}) as membrane-bound lipids active in the transport of sugar units to the polymeric cell wall has been clearly established.[13] The complexity inherent to the analysis of polyprenol mixtures may well have prevented the discovery of their tricyclic counterparts.

Added in proof: W. Herz and J. S. Prasad (*J. Org. Chem.*, 1984, **49**, 326) have just reported the synthesis of a C_{23} tricyclic terpene, assumed to be *ent*-(**2**), (R = Pr^n), which coeluted with the C_{23} terpene component of geological samples.

We thank the Société Nationale Elf-Aquitaine (SNEA) for financial support and supply of samples; Dr. J. Connan,

---

[§] The C_{22} homologue appears in significant amounts in some carbonate petroleums, in particular from the Aquitaine basin, France.

[¶] Mass spectral data for the synthetic tricyclic terpenes: C_{21}, m/z, 290(M^+, 22%), 275(18), 261(2), 234(3), 219(4), 205(5), 191(100), 137(29), 123(32), 109(32), 95(44), 81(40), 69(49), and 55(38); C_{22}, m/z, 304(M^+, 10), 289(14), 248(2), 233(2), 219(3), 191(100), 137(34), 123(55), 109(39), 95(46), 81(30), 69(24), and 55(10); C_{23}, m/z, 318(M^+, 18), 303(13), 262(3), 247(3), 233(2), 219(2), 205(7), 191(100), 137(31), 123(35), 109(30), 95(46), 81(42), 69(62), and 55(46); C_{24}, m/z, 332(M^+, 18), 317(16), 276(4), 261(2), 247(2), 219(3), 191(100), 137(23), 123(30), 109(25), 95(40), 81(34), 69(47), and 55(31); C_{25}, m/z, 346(M^+, 16), 331(14), 290(3), 261(3), 233(3), 191(100), 137(23), 123(31), 109(28), 95(44), 81(38), 69(50), and 55(37); C_{26}, m/z, 360(M^+, 18), 345(14), 331(3), 317(2), 275(3), 247(5), 191(100), 137(27), 123(32), 109(27), 95(53), 81(45), 69(62), and 55(60); C_{28}, m/z, 388(M^+, 19), 373(16), 332(4), 275(4), 261(2), 191(100), 137(22), 123(27), 109(26), 95(39), 81(30), 69(40), and 55(29); C_{29}, m/z, 402(M^+, 15), 387(13), 346(3), 317(2), 289(2), 191(100), 137(20), 123(25), 109(24), 95(38), 81(28), 69(41), and 55(27); C_{30}, m/z, 416(M^+, 14), 401(13), 360(3), 303(3), 261(2), 205(2), 191(100), 137(19), 123(22), 109(24), 95(39), 81(29), 69(41), and 55(25). The spectra were measured at 70 eV on a LKB 9000 S GC–MS, except for the C-22 compound for which a Kratos MS 80 GC-MS instrument was used. The 22*R* and 22*S* stereoisomers displayed nearly identical mass spectra.

SNEA(P), Pau, for his interest and helpful discussions; Professors A. J. Marsaioli, Universidade de Campinas, Brazil, and E. A. Ruveda, Universidad de Rosario, Argentina, for a sample of Copaiba oil; R. Hueber, G. Teller, and P. Wehrung for mass spectrometrical analysis. One of us (R. O.) thanks the Consejo Nacional de Ciencia y Tecnologia (CONACYT), Mexico, for financial support.

Received, 29th December 1983; Com. 1684

References

1 (a) W. K. Seifert and J. M. Moldowan, *Geochim. Cosmochim. Acta,* 1978, **42**, 77; (b) J. Connan, A. Restlé, and P. Albrecht, in 'Advances in Organic Geochemistry 1979,' eds. A. G. Douglas and J. R. Maxwell, Pergamon, Oxford, 1980, p. 1; (c) J. E. Zumberge, in 'Advances in Organic Geochemistry 1981,' eds. M. Bjøroy, P. Albrecht, C. Cornford, K. De Groot, G. Eglinton, E. Galimov, D. Leythaeuser, R. Pelet, J. Rullkotter, and G. Speers, Wiley, London, 1983, p. 738.

2 F. R. Aquino Neto, A. Restlé, J. Connan, P. Albrecht, and G. Ourisson, *Tetrahedron Lett.*, 1982, **23**, 2027.

3 F. R. Aquino Neto, J. M. Trendel, A. Restlé, J. Connan, and P. Albrecht, in ref. 1(c), p. 659.

4 C. M. Ekweozor and O. P. Strausz, *Tetrahedron Lett.*, 1982, **23**, 2711.

5 C. M. Ekweozor and O. P. Strausz, in ref 1(c), p. 746.

6 T. D. Cyr and O. P. Strausz, *J. Chem. Soc., Chem. Commun.*, 1983, 1028; J. D. Payzant, D. S. Montgomery, and O. P. Strausz, *Tetrahedron Lett.*, 1983, **24**, 651.

7 J. M. Moldowan, W. K. Seifert, and E. J. Gallegos, *Geochim. Cosmochim. Acta,* 1983, **47**, 1531.

8 E. D. McCarthy and M. Calvin, *Tetrahedron,* 1967, **23**, 2609.

9 B. J. Kimble, J. R. Maxwell, R. P. Philp, G. Eglinton, P. Albrecht, A. Ensminger, P. Arpino, and G. Ourisson, *Geochim. Cosmochim. Acta,* 1974, **38**, 1165.

10 A. Ensminger, A. Van Dorsselaer, C. Spyckerelle, P. Albrecht, and G. Ourisson, in 'Advances in Organic Geochemistry 1973,' eds. B. Tissot and F. Bienner, Technip, Paris, 1974, p. 245.

11 G. Ourisson, P. Albrecht, and M. Rohmer, *Trends Biochem. Sci.*, 1982, **7**, 236.

12 G. Ourisson, P. Albrecht, and M. Rohmer, *Pure Appl. Chem.*, 1979, **51**, 709.

13 F. W. Hemming, in 'Biochemistry of Lipids,' MTP International Review of Science, vol. 4, ed. T. W. Goodwin, Butterworths, London, 1973, p. 39.

14 S. Bory, D. Do Khac Manh, M. Fetizon, M. Kone, and A. Nguyen Trong, *Bull. Soc. Chim. Fr.*, 1975, 2347.

15 P. M. Imamura and E. A. Ruveda, *J. Org. Chem.*, 1980, **45**, 510.

16 W. Nagata and H. Itazaki, *Chem. Ind.* (*London*), 1964, 1194.

17 W. S. Wadsworth, Jr., and W. D. Emmons, *J. Am. Chem. Soc.*, 1961, **83**, 1733.

18 J. A. Profitt, D. S. Watt, and E. J. Corey, *J. Org. Chem.*, 1975, **40**, 127.

19 L. Crombie, P. Hemesley, and G. Pattenden, *J. Chem. Soc.* (*C*), 1969, 1016.