High-molecular-weight hydrocarbons: a new frontier in organic geochemistry

J. C. del Rio and R. P. Philp
Norman, OK, USA

For the past two decades organic geochemistry has been concerned mainly with compounds in the carbon number range up to $C_{40}$ as a result of various analytical limitations. The availability of gas chromatographic columns with upper temperature limits of 480°C, or even higher, has opened a new area for geochemical research, that of high-molecular-weight compounds above $C_{40}$. This paper reviews some recent advances in this field with particular reference to hydrocarbons and speculates on possible origins for some of the compounds.

Introduction

Traditionally, organic geochemistry has been concerned with the characterization of organic compounds present in the geosphere in the carbon number range $C_1$-$C_{40}$. Among them are the so-called biomarkers, organic compounds whose carbon skeletons can be linked unambiguously with known natural precursors. Biomarkers have been used extensively by organic geochemists to obtain information about source, thermal maturity, and depositional environments, of geologically-ancient sedimentary organic matter, and biodegradation of crude oils as well as information from Recent sediments related to paleoclimatic changes. One of the main reasons for the marked increase in the number of biomarkers that have been identified in the past few years has been improvements in analytical techniques, in particular gas chromatography–mass spectrometry (GC–MS). But in addition to the rapid increase in the number of biomarkers identified, there has also been a trend towards characterization of compounds with increasing molecular weight.

An earlier paper has reviewed the occurrence and significance of high-molecular-weight compounds isolated from sediments [1]. This report focused mainly on compounds such as triglycerides, alkylglycerols and wax esters, which can be hydrolyzed to produce smaller molecules. In the present report, however, the emphasis is on the presence of compounds having more than 40 carbon atoms and no ester (hydrolysable) or ether (degradable) linkages.

Historical overview

Over the past two decades, the carbon number range of different series of biomarkers has continued to extend steadily upwards. Thus, the first significant biomarkers that were recognized and used for geochemical purposes were compounds based on acyclic isoprenoid skeletons such as pristane ($C_{19}$) and phytane ($C_{20}$). These compounds were thought to be derived from a naturally occurring precursor, chlorophyll, although a number of alternative sources have been proposed subsequently. The range of the acyclic isoprenoids was extended with the discovery of isoprenoids in the $C_{35}$-$C_{45}$ region having both the regular 'head-to-head' and 'head-to-tail' structures [2,3]. Other examples of isoprenoid biomarkers detected at the current upper temperature limits for most conventional capillary columns are the $C_{40}$ tetraisoprenoids, like perhydro$^-$carotane—an important indicator for sediments deposited in lacustrine environments—and the $C_{40}$ biphytanyl residues of the alkylglycerol ethers, found as major lipid constituents of archaebacteria and reported in the bitumen from sediments and in some petroleum [4–7].

A similar pattern can be found with the tricyclic terpanes where the molecular weight range of known compounds has been extended gradually to $C_{48}$ and with solanesol, $C_{45}$; polypropenol, suggested as a possible precursor for this series [8]. The first members of this series in the $C_{20}$-$C_{23}$ region were identified in the Athabasca tar sands [9]. Pentacyclic triterpanes with hopanoid skeletons are an important class of triterpenoid biomarkers ubiquitous in sediments and fossil fuels. The hopanoid hydrocarbons, are generally found in the range $C_{27}$-$C_{33}$ but this has been extended up to the $C_{40}$ homologues [10].

Organically bound sulphur in fossil fuels and related sedimentary organic matter is present in both low-molecular-weight organic sulphur compounds (OSC) and in complex macromolecules (kerogens, asphaltenes and resins). The range of organic sulphur compounds identified has also in-
increased, for example, with the identification of thiophene hopanoids up to C_{90}, linear and isoprenoide thiophenes up to C_{40} and tetracyclic terpenoid sulphides up to C_{60} [11]. The investigation of other classes of biomarkers, such as chlorins and porphyrins, has shown a similar trend towards the discovery of novel high-molecular-weight compounds [12,13].

In the last two years there has been an increase in reports concerned with the detection and identification of high-molecular-weight hydrocarbons. Van Aarssen and de Leeuw [14] analyzed crude oils from South East Asia for cyclic terpenoids originating from polycadinane, a (bio)polymer present in dammar resins. They found, among cadinanes, murolanes and bicadinanes, two hydrocarbons with molecular formulae C_{45}H_{76} and C_{49}H_{80}, thought to be trimeric cadinanes. In the aromatic fraction they encountered a series of C_{42} compounds, considered to be di-aromatic tricadinanes (Fig. 1). These components result from depolymerization of the dammar resin polymer and subsequent hydrogenation or dehydrogenation and can be used as markers for angiosperm plants such as the Dipterocarpaceae [14].

Recently del Rio and Philp [15] analysed waxes from an oil pipeline by high temperature gas chromatography (HTGC), and observed a very high concentration of high-molecular-weight hydrocarbons (up to at least C_{60}) in the waxes, that are not present in the oils collected at the well-head (Fig. 2). The waxes, although dominated by n-alkanes, also contained complex mixtures of branched/cyclic hydrocarbons, mainly alkylcyclohexanes and mid-chain methyl alkanes, probably formed by condensation reactions of fatty acids [16]. Similar results have also been reported by other authors [17–20] who identified series of co-eluting monocyclic and acyclic saturated hydrocarbons extending to the C_{60} molecular weight range in a variety of crude oils. These compounds showed odd carbon number preference for marine-sourced oils and an even number preference for lacustrine-sourced oils.

Kohnen et al. [21] using HTGC, reported the occurrence of long chain (C_{36}–C_{54}) 3,4-dialkyl thiophenes in immature sediments, and in some samples these compounds were the major alkylthiophenes present. It was proposed that mid-chain dimethyl hydrocarbons with two methylene double bonds were suitable precursors for the incorporation of sulphur, although no natural precursors
have been reported in organisms and/or sediments with this type of carbon skeleton.

The higher molecular weight hydrocarbons in oils are more resistant to chemical and/or biological degradation since it is known that biodegradation initially affects the lower molecular weight components. These compounds, or their precursors, may also be incorporated into the asphaltene or kerogen fractions. Hence, in addition to investigating the high-molecular-weight components in the soluble fractions, it is also important to investigate the high-molecular-weight degradation products produced by pyrolytic degradation of kerogen and asphaltenes. The presence of hydrocarbons up to $C_{30}$ has already been reported in the pyrolyzates of several asphaltenes [15,22] (Fig. 3). However, it is still not absolutely clear whether these higher number hydrocarbons are actually pyrolysis products from the asphaltenes or represent the release of unbound components trapped in the asphaltenes. Asphaltenes, by definition, are a solubility class precipitated from crude oils and sediment extracts by the addition of a large excess of a low-boiling liquid hydrocarbon, commonly $n$-pentane. Under these conditions high-molecular-weight hydrocarbons may precipitate and become part of the actual asphaltene fraction. In previous pyrolysis studies performance capabilities of the chromatographic columns have limited the pyrolysis products that could be analyzed to approximately $C_{40}$ [23]. However, the combination of high temperature capillary columns and the extended mass ranges of mass spectrometers should provide an excellent means for investigating these high-molecular-weight pyrolysis products.

Despite the advances made in the past few years, the study of higher molecular weight hydrocarbons has been largely overlooked for two main reasons. The first is the lack of appropriate techniques, since traditional GC methods are limited to species with sufficient volatility. The second, particularly in the case of the oils, is that the high-molecular-weight hydrocarbons are often not present, or in very low concentrations in the oils collected at the well head. Rather, these compounds may precipitate due to their low solubility and mobility at surface temperatures and pressures. In all of these examples, improvement of instrumentation, major advances in GC column technology and a greater understanding of the origin of these compounds have combined to greatly extend the molecular weight range of the biomark-

Asphaltenes oil FRL 9169

Fig. 3. High temperature capillary gas chromatogram of the products obtained after pyrolysis of the asphaltenes isolated from a crude oil.
ers that are currently being discovered and utilized in many areas of geochemistry. This review summarizes these analytical and geochemical advances.

**Recent developments in analytical techniques**

New techniques have become available recently to deal with the analysis of high-molecular-weight hydrocarbons (>C\text{40}), including HTGC and supercritical fluid chromatography (SFC). GC capillary columns (fused silica, aluminum and stainless-steel capillary columns) contain thin films of thermally stable, bonded, and crosslinked stationary phases which can be maintained isothermally at 400–425°C and temperature programmed to 425–440°C. Under these conditions, alkanes with carbon numbers in the C\text{90}–C\text{100} range can be analyzed \[24,25\]. HTGC has also been used to analyze free base porphyrins and metal porphyrin complexes in oil shale extracts \[26,27\].

Capillary SFC is another rapidly developing method for the separation of compounds that lack sufficient volatility to be separated by GC and its use to separate high-molecular-weight compounds has been demonstrated by several authors \[28–34\]. SFC is now being recognized as an important method for analytical separation, since the physical-chemical properties of supercritical fluids impart significant chromatographic advantages for the analysis of thermally labile and higher molecular weight compounds. In addition, the use of small diameter capillary columns (25–100 μm I.D.) for SFC, which have negligible pressure drops, allows exploitation of the full potential of pressure programming methods to obtain high resolution separation, approaching that achievable with capillary GC.

The coupling of SFC or HTGC with MS offers the potential to provide high resolution separation of very high-molecular-weight compounds with selective detection \[24,25,30,33,35\], and are the appropriate techniques required to deal with the characterization of the high-molecular-weight biomarkers. Even simultaneous coupling of SFC and HTGC with a single quadrupole mass spectrometer has been described in the literature \[36\]. Recently, HTGC-field ionization mass spectrometry (FIMS) has also been successfully applied to the analysis of porphyrins and complex mixtures of high-molecular-weight saturated and aromatic hydrocarbons \[17–19\]. The FIMS method permits the identification of molecular weights (and thus ring number) of high-molecular-weight hydrocarbons that exhibit no molecular ions in MS studies employing other ionization methods. Hawthorne and Miller \[31\], using SFC–MS, analyzed several commercial waxes, including a microcrystalline wax, consisting of n-alkanes ranging up to C\text{35}, and synthetic Fischer-Tropsch wax, made of alkanes up to C\text{100}. The potential of SFC–MS will continue to grow as SFC techniques are extended by the introduction of new mobile and stationary phases and as improved MS detectors for higher molecular weight compounds (>1000) become available.

**Possible origins and methods of formation of high-molecular-weight hydrocarbons**

**Natural occurrence of high-molecular-weight biochemical precursors**

The presence of compounds in living organisms having more than 40 carbon atoms has been widely reported. Polysoprenoid alcohols (polyprenols) is one of the simplest groups of naturally occurring terpanes that extend up to more than C\text{120} (Fig. 4) and they have already been suggested as possible precursors for isoprenoids and tricyclic terpanes beyond C\text{40} \[3,8,9\]. They are important intermediates in the biosynthesis of polysaccharides and consist of a long chain of isoprene units (6 to 24) linked head-to-tail with a hydroxyl group at one end. Chojnacki and Vogtman \[37\] reported that many angiosperms accumulated polyprenols larger than compounds previously known to be present in the leaves of these plants. The presence of C\text{120} polyprenols in conifers and the occurrence of more than one polyprenol family in pine needles has been documented \[38\], as well as their occurrence in microorganisms \[39\].

Carotenoids are another class of compounds of geochemical interest where bacterial C\text{45} and C\text{40} carotenoid skeletons had been reported \[40\]. The C\text{50}-skeletal bacterioruberin from halophilic bacteria and the C\text{50} sarcinaxanthin (Fig. 5) are two examples of these higher molecular weight compounds. Mycolic acids, β-hydroxy fatty acids substituted in the α-position (Fig. 5) is another class of high-molecular-weight biochemical compounds and very characteristic of Actynomicetes \[41\].

High-molecular-weight compounds above C\text{40}, or their diagenetic products, have not been studied in the fossil record by organic geochemists in the same detail as those compounds below C\text{40}. There is no compelling reason why at least some of these compounds, or their diagenetic products,
should not be preserved in the sedimentary record and detected if the appropriate analytical techniques are used. The alcohols may be converted into high-molecular-weight isoprenoid alkanes by geochemical processes such as reduction of the functional group followed by thermal cracking similar to their low-molecular-weight analogues (i.e. phytol).

**Formation in sediments**

One of the reasons high-molecular-weight compounds may not always be observed in the bitumen fraction of sediments and fossil fuels is that they, or their precursors, may directly be incorporated into the asphaltene and/or kerogen fractions. Degradation of the asphaltenes at high temperatures and/or pressures in reservoirs will produce hydrocarbons (low and high molecular weight) which subsequently dissolve in the whole oil. The presence of higher molecular weight hydrocarbons in the flash pyrolysis–HTGC of some asphaltenes [15,16] supports this hypothesis.

Previously published papers strongly support the idea that some of these compounds are di- and trimerization products of lower molecular weight precursors. For example de Leeuw et al. [42] obtained an aromatic dimerization product and unidentified trimerization products from phytol, after heating phytol products in presence of clay minerals. Rubinstein and Strausz [43], demonstrated the formation of cyclohexyl structures by heating fatty acids in the presence of clay minerals, and also reported the formation of some dimerization products. We repeated the same experiments in our laboratories, analyzed the products by HTGC and observed the formation of trimeric products [22]. The structures of the compounds produced (mid-chain methyl-branched alkanes and long-chain dialkycyclohexanes) were very similar to the high-molecular-weight hydrocarbons previously observed in geochemical materials. Hence naturally occurring long-chain fatty acids may be precursors for some of the high-
molecular-weight hydrocarbons occurring in geological samples. Furthermore, any organic compound with appropriate functional groups could oligomerize under appropriate conditions to form higher molecular weight compounds. Hence tricadinanes found in some crude oils [14] may be oligomerization products of cadinane monomers.

Fischer-Tropsch synthesis, a well-known industrial process whereby millions of tons of hydrocarbon oil resembling petroleum are produced annually from carbon monoxide or carbon dioxide and hydrogen gases reacting on an iron-oxide catalyst, makes it possible to generate high-molecular-weight hydrocarbons. A somewhat speculative hypothesis of petroleum generation based on Fischer-Tropsch synthesis has been proposed [44]. A great deal of additional work is needed to prove or reject this hypothesis as a means of formation of high-molecular-weight hydrocarbons. The processes described above may also occur simultaneously and are not mutually exclusive and hence these high-molecular-weight compounds might have multiple sources.

**Conclusions**

The continuing study of high-molecular-weight hydrocarbons (>C40) will provide additional insights into the origin and types of organic source materials responsible for various types of oils as well as fossil resins and bitumens. Major advances can be expected in the next few years because of continued improvements in analytical techniques that will make it easier to identify high-molecular-weight hydrocarbons at the detailed molecular-structure level. These advances will open up a new frontier area of biomarker geochemistry and will help to clarify our understanding on the preservation of biogenic molecules as well as the diagenetic pathways that lead to the formation of fossil fuels. As a result, knowledge about petroleum composition and genesis will increase greatly.

Results from the study of high-molecular-weight compounds in geological materials will be potentially useful to extend geochemical theories based on results already derived from compounds in the range C1–C40 and will provide us with additional tools for interpreting the geological record. It would appear that we are now on the verge of a new era and about to move into the analysis of components above C40, probably up to C100 and possibly even higher molecular weights.
References


Dr. J. C. del Rio and R. P. Philp are at the School of Geology and Geophysics, University of Oklahoma, Norman, OK 73019, USA. The permanent address of Dr. del Rio is: Instituto de Recursos Naturales y Agrobiologia, C.S.I.C., P.O. Box 1052, 41080 Sevilla, Spain.