Kirk-Othmer Encyclopedia of Chemical Technology.

PETROLEUM, COMPOSITION

Petroleum, literally rock oil, describes a myriad of hydrocarbon-rich fluids present in source rocks and accumulated in subterranean reservoirs (see Hydrocarbons). Petroleum can include three phases: gaseous (natural gas), liquid (crude oil), and solid or semisolid (bitumens, asphalt (qv), tars, and pitches) (1) (see Coal; gas, natural; Tar and pitch). The molecular composition of the liquid portion of petroleum contributing to the crude oil properties and behavior is discussed herein. Crude oils vary dramatically in color, odor, and flow properties. These properties often reflect the origin of the crude. Historically, physical properties such as boiling point, density (gravity), odor, and viscosity have been used to classify oils (2–4). Crude oils may be called light or heavy in reference to relative density (or specific gravity). Light crude oils are rich in low boiling and paraffinic hydrocarbons; heavy crude oils contain greater amounts of high boiling and asphalt-like molecules. The heavy oils tend to be more viscous, higher boiling, more aromatic, and contain larger amounts of heteroatoms. Likewise, odor is used to distinguish between sweet or low sulfur, and sour or high sulfur, crude oils.

Petroleum is thought to be derived from a variety of living organisms buried with sediments in previous geological eras. A small fraction of these organisms were trapped in oxygen-deficient (or reducing) environments where they escaped complete oxidation to carbon dioxide. Over tens or hundreds of millions of years, the residual organic material was subjected to a complex series of chemical changes known as diagenesis and catagenesis (2, 3). In diagenesis, which occurs below 50°C, the organics undergo microbial action and some chemical reactions, resulting in dehydration, condensation, cyclization, and polymerization. During catagenesis, which occurs under a thermal stress of 50–200°C, the organics react with the surroundings by a combination of thermocatalytic cracking, decarboxylation, and hydrogen disproportionation to form petroleum in the sedimentary rocks. In the vast majority of cases the petroleum is not found where the precursors were laid down, but in reservoirs where accumulation occurs after migration from the source rocks through geologic strata (3, 4) (see Petroleum, origin of petroleum).

The distribution of biomarker isomers, molecules that retain the basic carbon skeletons of biological compounds from living organisms, serves not only as a set of fingerprints for oil–oil and oil–source correlation (to relate the source and reservoir for exploration), but also to give geochemical information on organic source input (marine, lacustrine, or terrigenous source), age, maturity, depositional environment (clay or carbonate, oxygen levels, salinity, etc), and alteration (water washing, biodegradation, etc) (5, 6).

Knowledge of the composition of petroleum allows the refiner to optimize conversion of raw petroleum into high value products (4). Originally, petroleum was distilled and sold as fractions, primarily for use in illumination and lubrication (see Lubrication and lubricants). As of this writing (ca 1995), crude oil is sold in the form of gasoline (see Gasoline and other motor fuels), solvents (see Solvents, industrial), diesel and jet fuel (see Aviation and other gas turbine fuels), heating oil, lubricant oils, and asphalts, or it is converted to petrochemical feedstocks (see Feedstocks, petrochemicals) such as ethylene(qv), propylene(qv), the butenes, butadiene, and isoprene. Modern refining uses a sophisticated combination of heat, catalyst, and hydrogen(qv). Conversion processes include coking, hydrocracking, and catalytic cracking to break large molecules into smaller fractions; hydrotreating to reduce heteroatoms and aromatics, thereby creating environmentally acceptable products;

Table 1. Compounds Found in Petroleum Crude Oils[a]

Compound	CAS Registry Number	Molecular formula	Structure number
	Saturates		
n-octane	[111-65-9]	C_8H_{18}	(**1**)
2-methyloctane	[3221-61-2]	C_9H_{20}	(**2**)
propylcyclohexane	[1678-92-8]	C_9H_{18}	(**3**)
	Aromatics		
n-butylbenzene	[104-51-8]	$C_{10}H_{14}$	(**4**)
1-methylnaphthalene	[90-12-0]	$C_{11}H_{10}$	(**5**)
9-methylphenanthrene	[883-20-5]	$C_{15}H_{12}$	(**6**)
	Sulfur compounds		
propyl mercaptan	[107-03-9]	C_3H_8S	(**7**)
methyl propyl sulfide	[3877-15-4]	$C_4H_{10}S$	(**8**)
dibenzothiophene	[132-65-0]	$C_{12}H_8S$	(**9**)
	Oxygen compounds		
phenol	[108-95-2]	C_6H_6O	(**10**)
2-phenanthrene carboxylic acid	[40452-20-8]	$C_{15}H_{10}O_2$	(**11**)
cyclohexyl carboxylic acid	[98-89-5]	$C_7H_{12}O_2$	(**12**)
	Nitrogen compounds		
quinoline	[91-22-5]	C_9H_7N	(**13**)
carbazole	[86-74-8]	$C_{12}H_9N$	(**14**)
2(1*H*)-quinolinone	[59-31-4]	C_9H_7NO	(**15**)

[a] See Fig. 1.

and isomerization and reforming to rearrange molecules to those having high value, eg, gasolines of high octane number.

A knowledge of the molecular composition of a petroleum also allows environmentalists to consider the biological impact of environmental exposure. Increasingly, petroleum is being produced in and transported from remote areas of the world to refineries located closer to markets. Although only a minuscule fraction of that oil is released into the environment, the sheer volume involved has the potential for environmental damage. Molecular composition can not only identify the sources of contamination but also aids in understanding the fate and effects of the potentially hazardous components (7).

Crude oils contain an extremely wide range of organic functionality and molecular size. The variety is so great that a complete compound-by-compound description for even a single crude oil is not likely. The composite molecular composition of petroleum can, however, be described in terms of three classes of compounds: saturates, aromatics, and compounds bearing the heteroatoms sulfur, oxygen, or nitrogen. Within each of these classes there are several families of related compounds. Table 1 lists some of the compounds typically found in petroleum crude oils. Structures are shown in Figure 1. The saturates include normal alkanes, branched alkanes, and cycloalkanes called paraffins, isoparaffins, and naphthenes, respectively (see Petroleum, nomenclature in the petroleum industry). Alkenes (olefins) are rare to the extent of being considered an oddity. Aromatics range from benzene to multiple fused-ring analogues such as naphthalene, phenanthrene, etc. Sulfur is found in both polar (thiol) and nonpolar (thioether and thiophenic) forms (8). The nitrogen and oxygen compounds are more likely to be found in polar compounds such as pyridines, pyrroles, phenols, carboxylates, amides, etc, than in nonpolar ones such as ethers (9). The distribution and characteristics of these molecular species account for the rich variety of crude oils.

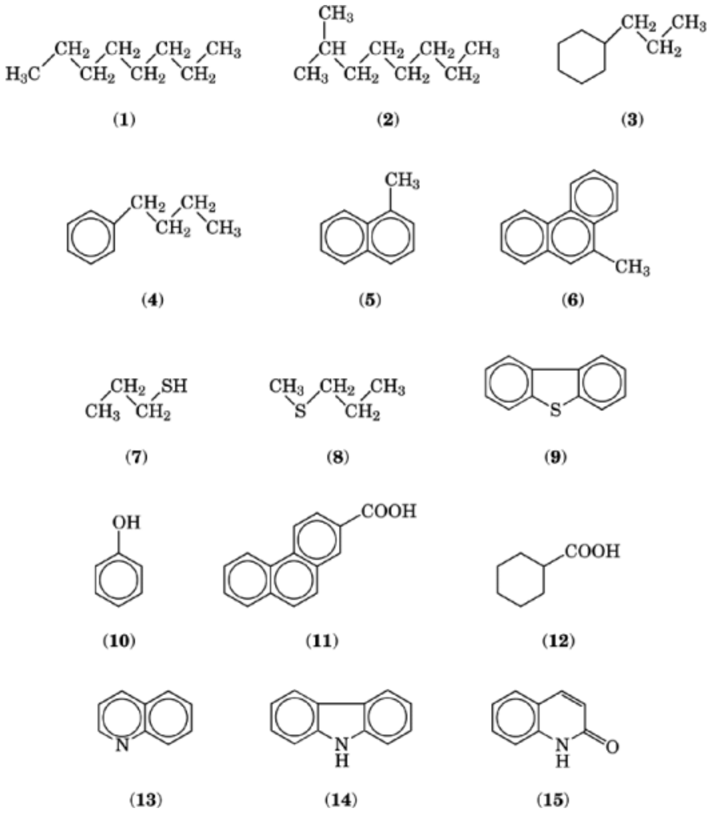


Fig. 1. Structures of compounds in petroleum crude oils. See Table 1.

1. Elemental Composition

On an atomic basis, H/C ratios range from 1.5–2.0. The range of elemental composition of crude oil (3) may be given as follows:

Element	Composition range, wt %
carbon	84–87
hydrogen	11–14
sulfur	$<0.1 - 8$
oxygen	$<0.1 - 1.8$
nitrogen	$<0.1 - 1.6$

Nickel and vanadium are also generally present from a trace amount up to 1000 ppm. The value of using sulfur for petroleum classification can be clearly seen. Whereas the range of hydrogen and carbon are quite narrow, sulfur, the principal heteroatom in crude oil, varies significantly. The ranges for the other two heteroatoms (O and N) are also fairly wide, however in most crude oils these elements are nearly an order of magnitude lower than the sulfur level (Fig. 2). For example, the average level of nitrogen in >9000 crude oils is 0.094%, the average sulfur content is 0.65% (2). Although $>90\%$ of crude oils have nitrogen levels $<0.2\%$, some crude oils have $>$ 1.5% nitrogen. As noted in Figure 2**b**, the nitrogen concentrates into the highest

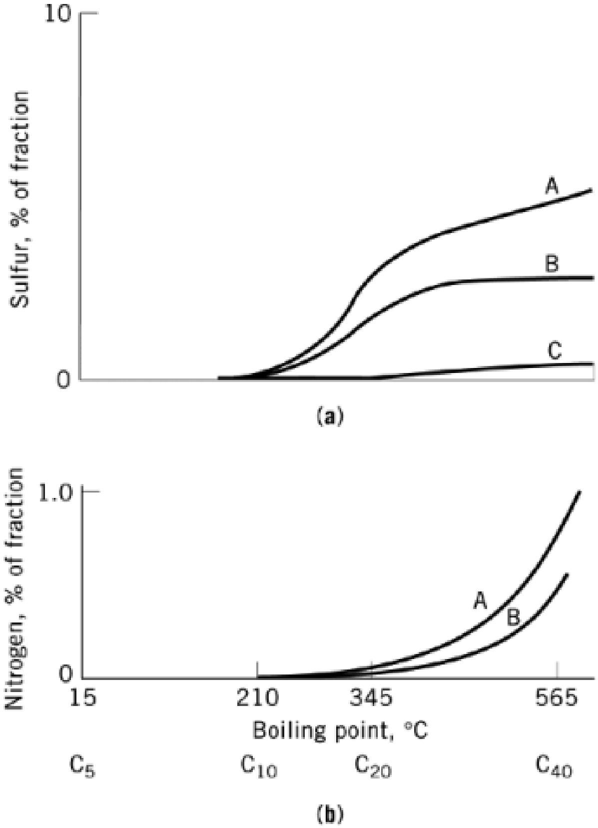


Fig. 2. Distribution of compounds as a function of crude oil boiling point: (**a**) sulfur where A, B, and C represent high ($>2\%$), medium (ca 1.5%), and low ($<0.1\%$) sulfur, respectively, and (**b**) nitrogen, where A and B represent high (ca 0.5%) and low ($<0.1\%$) nitrogen, respectively.

boiling fractions. The trace metals Ni and V are generally orders of magnitude higher than other metals in petroleum, unless it is contaminated with coproduced brine salts (Na^+, Mg^{2+}, Ca^{2+}, Cl^-) or iron corrosion products gathered in transportation.

2. Molecular Classes

The molecules in crude oil include several basic structural types (Table 1, Fig. 1). Because they may contain from 1 to 100+ carbon atoms and may occur in combination, the statistical potential for isomeric structures is staggering. For example, whereas there are just 75 possible paraffinic structures for C_{10}, there are $>10^5$ isomers for C_{20}. A few structures tend to dominate the distributions of each isomer group, however.

The inclusion of naphthene and other aromatic rings introduces two additional dimensions, increasing the number of hydrocarbon isomers even further. A three-dimensional array in which the molecules could be described in terms of the number of aromatic rings, the number of naphthenic rings, and the number of carbons in alkyl side chains has been proposed (9). Conceptually, this amounts to describing a three-dimensional molecular mountain for hydrocarbons. There is also the potential of constructing similar mountains for heterocyclics.

The two-dimensional image in the naphthenic and aromatic dimensions has been projected using sidebars to indicate the variation in alkyl substituents among crude oils (2).

Molecular characterization of a whole oil is beyond the capability of most analytical techniques. Distillation (qv), however, can separate petroleum into molecular weight fractions that simplify the task. Pioneering work with this approach, sponsored by the American Petroleum Institute (API) starting in 1925, has led to the identification of hundreds of individual compounds in distillation fractions of a single crude oil (10). More recently, developments in chromatography (qv) allowed oils to be fractionated by polarity as a second dimension. Under API sponsorship, the U.S. Bureau of Mines extended separations and measurement techniques to heavier fractions (11). At the same time, individual compounds have been isolated and quantified from increasingly higher boiling fractions (12–15). Techniques have been developed that use combinations of classical open-column adsorption chromatography, gel permeation chromatography, and ion-exchange (qv) separations to isolate fractions in which compounds could be identified by mass spectrometry (qv).

Whereas neither distillation nor chromatography achieves perfect separations among groups, the fractions generated are amenable to molecular characterization (14, 15). An elegant argument has been made for the use of distillation for primary separation which is combined with a solubility step to achieve the atmospheric equivalent boiling point (AEBP) scale that covers room temperature to 1370°C (16). Because the distillations simplify analytical complexity, the bulk of available molecular compositional data on petroleum has been generated on sets of fractions defined by boiling points as gases (C_1–C_4), naphtha (initial bp –210°C), mid-distillate (210–345°C), vacuum gas oil (345–565°C), and vacuum residuum (>565°C). The initial boiling point (ibp) of the naphtha depends on the amount of C_1–C_5 hydrocarbons dissolved in the oil. For characterization purposes, naphtha is usually de-pentanized so that the ibp is about 32°C. The material that boils above 345°C is referred to as atmospheric resid and is further distilled under vacuum into vacuum gas oil (VGO) and vacuum resid. Whereas the AEBP of the vacuum resid is >565°C, the actual temperature of distillation does not exceed 345°C to avoid thermal decomposition.

Crude distillations yield different quantities in each fraction. About the same amounts are distilled into the middle distillate and vacuum gas oil from conventional crude oils. More naphtha is distilled from light crude oils and more vacuum residuum is obtained from heavy crude oils (Fig. 3). The typical distribution of classes of petroleum compounds shows a significant shift with boiling point (Fig. 4). Whereas the lower boiling fractions are dominated by nonpolar saturated hydrocarbons that exist in limited isomeric forms, the higher boiling fractions increasingly contain a larger variety of classes, that have, in turn, an increasing number of possible isomers. As the boiling point increases, aromatic ring structures build in, first as naked rings, then more and more as rings having attached side-chain and naphthene ring carbons. Polar compounds, typically those having O and N functionality, that appear as trace impurities in the lower boiling fractions gradually become significant components in the higher boiling fractions. This is confirmed by the distribution of S and N in petroleum against boiling point (Fig. 2). The S, not including H_2S and the light sulfur compounds such as mercaptans and sulfides, present in petroleum gases, is more widely distributed than the nitrogen that concentrates in the highest boiling fraction. Not shown is the subtle decrease that occurs in H/C ratio with increasing boiling point reflecting the increasing number of aromatic ring types at higher boiling point. The metals, nitrogen, and oxygen are predominantly found in the higher boiling fractions rich in polars.

2.1. Analytical Approaches

Different analytical techniques have been applied to each fraction to determine its molecular composition. As the molecular weight increases, complexity increasingly shifts the level of analytical detail from quantification of most individual species in the naphtha to average molecular descriptions in the vacuum residuum. For the naphtha, classical techniques allow the isolation and identification of individual compounds by physical properties. Gas chromatographic (gc) resolution allows almost every compound having less than eight carbon

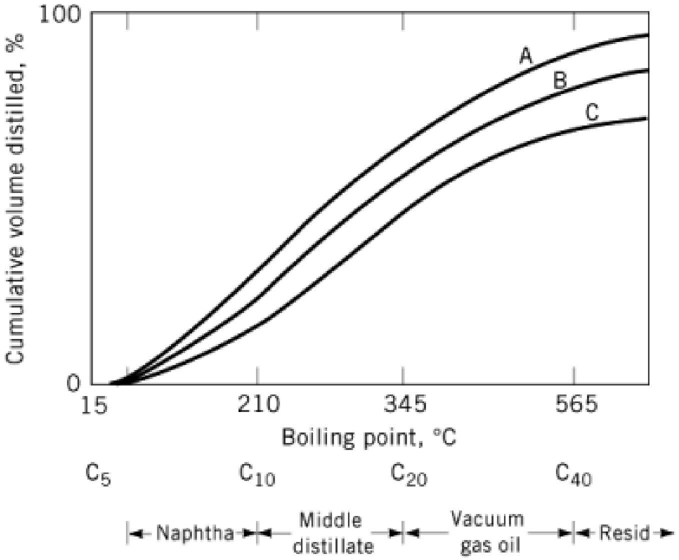


Fig. 3. Cumulative volume distilled as a function of boiling point from A, light; B, intermediate; and C, heavy crude oils (not including condensates).

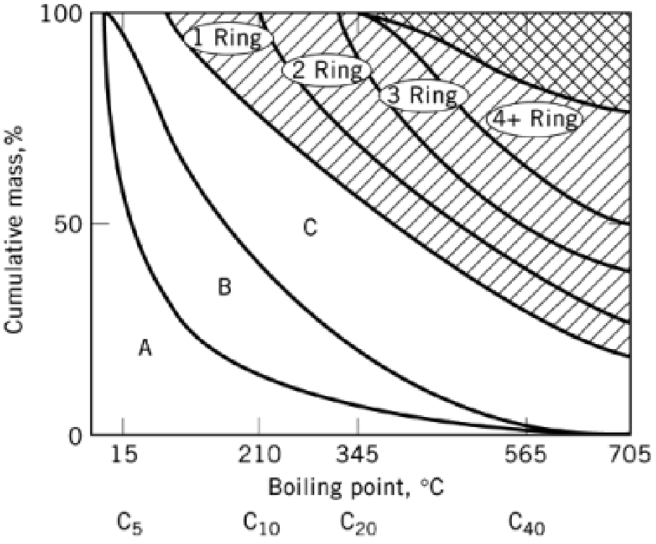


Fig. 4. Distribution of compound classes in crude oils as a function of boiling point. Region A represents normal paraffins; B, isoparaffins; C, naphthenes; □ the region of alkyl and napthenic aromatics; and▩ the region of polars.

atoms to be measured separately. The combination of gc with mass spectrometry (gc/ms) can be used for quantitation purposes when compounds are not well-resolved by gc.

For the mid-distillates and vacuum gas oils (VGOs), class isolation and measurement techniques allow speciation of many compounds. In particular, multidimensional instrumental techniques have been applied. Capillary gc quantifies even closely related isomers. Element-selective detectors for gc are used to identify N, S, and O heterocompounds; gc/ms techniques are used to identify and quantify individual compounds or families of compounds. Characterization of families of compounds having similar degrees of unsaturation, expressed by z-number as in $C_nH_{2n+z}X$, where X represents heteroatoms, can be done by ms quantification (17). The combination of high performance liquid chromatography (hplc) and high resolution/low voltage ms has also been used to resolve overlaps between aromatic hydrocarbons and sulfur analogues as well as among isomers of alkyl aromatics and naphthenoaromatics (17). For VGOs ms techniques give semiquantitative results owing to the lack of reference compounds for calibrating relative sensitivities. Alternative techniques such as ultraviolet detection may be needed to distinguish among structural isomers (18, 19).

The combination of chromatographic isolation followed by ms measurements has been extended well above 565°C into fractions of the vacuum resid (20–22). For fractions not amenable to that approach, techniques that provide average functionality data have been applied to describe petroleum composition (20). Established techniques such as titration for acidic and basic functionalities and for sulfur types, ultraviolet (uv) spectroscopy for aromaticity, and infrared (ir) spectroscopy for dipolar functionality have been supplemented with alternative instrumental techniques to expand the average database. These include nuclear magnetic resonance (nmr) for aromaticity, x-ray photoelectron spectroscopy (xps) for chemical bonding, and extended x-ray absorption fine structure (exafs) for atomic coordination environment, as well as improved Fourier-transform ir and Raman techniques for functionality (see Infrared technology and raman spectroscopy; Magnetic spin resonance; Spectroscopy). The individual average techniques do not provide molecular composition (23). On the other hand, these techniques can provide an average composition when used in combination. For those portions of the vacuum residua that are both nonvolatile and insoluble and hence not accessible to molecular speciation techniques, the alternative techniques are helpful.

2.2. Petroleum Gases and Naphtha

Methane is the main hydrocarbon component of petroleum gases. Lesser amounts of ethane, propane, butane, isobutane, and some C_4+ light hydrocarbons also exist. Other gases such as hydrogen, carbon dioxide, hydrogen sulfide, and carbonyl sulfide are also present.

The naphtha fraction is dominated by saturates having lesser amounts of mono- and diaromatics (Table 2, Fig. 4). Whereas naphtha (ibp to 210°C) covers the boiling range of gasoline, most raw petroleum naphtha molecules have a low octane number and most raw naphtha is processed further, to be combined with other process naphthas and additives to formulate commercial gasoline.

Table 2. Compounds Found in Petroleum Naphtha[a]

Compound	CAS Registry	Molecular formula	Structure number
n-heptane	[142-82-5]	C_7H_{16}	(**16**)
2-methylheptane	[592-27-8]	C_8H_{18}	(**17**)
methylcyclopentane	[96-37-7]	C_6H_{12}	(**18**)
ethylcylohexane	[1678-91-7]	C_8H_{16}	(**19**)
1,3-dimethylbenzene (m-xylene)	[108-38-3]	C_8H_{10}	(**20**)
indan	[496-11-7]	C_9H_{10}	(**21**)
naphthalene	[91-20-3]	$C_{10}H_8$	(**22**)
tetrahydronaphthalene (tetralin)	[119-64-2]	$C_{10}H_{12}$	(**23**)
decahydronaphthalene (decalin)	[91-17-8]	$C_{10}H_{18}$	(**24**)
methyl mercaptan	[74-93-1]	CH_4S	(**25**)
3-methylthiacyclohexane	[5258-50-4]	$C_6H_{12}S$	(**26**)
3-(methylthio)pentane	[57093-84-2]	$C_6H_{14}S$	(**27**)

[a] See Fig. 5.

Within the saturates in petroleum gases and naphtha, except for a few highly branched components in the C_8–C_{10} range, every possible paraffin from methane to normal decane (n-C_{10}) is present. Depending on the source, one of the low boiling paraffins may be the most abundant compound in a crude oil, sometimes reaching a concentration of several percent. The isoparaffins begin at C_4. Isobutane is the only isomer of n-butane. The number of isomers grows rapidly with carbon number so that there are 74 isomers of C_{10}. Some of the individual 2-methyl isoparaffins may be present in concentrations $>1\%$. Using instrumental techniques every possible isoparaffin in petroleum up to C_8 has been identified. Although complete resolution of all isomers $>C_8$ is beyond analytical capabilities, many of these compounds have also been identified.

In addition to the aliphatic (chain) molecules, the saturates contain cycloalkanes, called naphthenes, having mainly five or six carbons in the ring (Fig. 5). Methyl derivatives of cyclopentane and cyclohexane are commonly found in greater quantity than the parent unsubstituted structures and can be present at levels above 2% (2). Fused-ring dicycloalkanes such as decahydronaphthalenes (decalins) and hexahydroindans are also common, but nonfused bicylic naphthenes, eg, cyclohexyl cyclohexane, are not.

The numerous aromatics in petroleum naphtha begin with benzene, but the C_1–C_3 alkylated derivatives of benzene generally are present in larger amounts. Toluene concentrations may reach nearly 2%; the combined xylene isomers exceed 1%; benzene, however, rarely exceeds 1% of a whole crude oil (2). Although present in lesser concentrations, each of the alkyl benzene homologues through the 20 isomeric C_4 alkyl benzenes have been isolated from Ponca City crude, as have several of the C_5-derivatives (24). Benzenes having fused cycloparaffin rings (naphthenoaromatics) such as indan and tetralin have been isolated along with a number of their methyl derivatives. Naphthalene is included in the naphtha, whereas the 1- and 2-methyl naphthalenes and higher homologues of fused two-ring aromatics appear in the mid-distillate fraction.

Sulfur is the only heteroatom to be found in naphtha, and only at trace levels (Fig. 2). A systematic study of organic sulfur compounds in Wasson Texas crude, and to a lesser extent three other crude oils, culminated in the identification of over 200 compounds, most of which were in the naphtha fraction (24). However, the total of these compounds accounted for less than 1% of the sulfur in the whole crude. In sour petroleum naphthas, 50–70% of the sulfur can be found in the form of mercaptans, ie, thiols. Over 40 individual thiols have been identified, including all the isomeric C_1 to C_6 compounds plus some C_7 and C_8 isomers and thiophenol (25). In sweet naphthas, the sulfur is distributed between sulfides, ie, thio-ethers, and thiophenes. In these cases the sulfides may constitute more than 50% of the sulfur compounds in the form of both linear (alkyl sulfides) and five- or six-ring cyclic (thiacyclane) structures. Sulfur structure distribution mimics the hydrocarbons; ie, naphthenic oils having high cycloalkanes have high thiacyclane content, etc. Typical alkyl thiophenes in naphtha have multiple short side chains or exist as naphthenothiophenes (25). Though rare, methyl and ethyl disulfides have been confirmed to be present in some crude oils (26).

2.3. Mid-Distillates

As is indicated in Figure 4, saturates remain the primary component in the mid-distillate fraction of petroleum, but aromatics, which include simple compounds having up to three aromatic rings, and heterocyclics are present and represent a larger portion of the total. Some raw middle distillates are used directly as kerosenes, jet fuels, and diesel fuels; others are cracked and hydroprocessed before use. Some compounds found in middle distillates are listed in Table 3. Within the saturates, the concentration of n-paraffins generally decreases regularly from C_{11} to C_{20}. The 2-methyl analogues are sufficiently unique to be seen as distinct peaks in gc analyses of mid-distillates. Few additional isoparaffins have been identified. However, the two isoprenoids (**30**) and (**31**) (see Fig. 6) are generally present in crude oils in sufficient concentration to be seen as irregular peaks alongside the n-C_{17} and n-C_{18} peaks in a gas chromatogram. These isoprenoids, believed to arise primarily as fragments of ancient chlorophyll [1406-65-1] have relevance, as simple biomarkers, to the genesis of petroleum. The distribution of pristane and phytane relative to the neighboring n-C_{17} and n-C_{18} peaks has been used to aid in the identification of crude oils and to detect the onset of biodegradation (3). The ratio of pristane to phytane

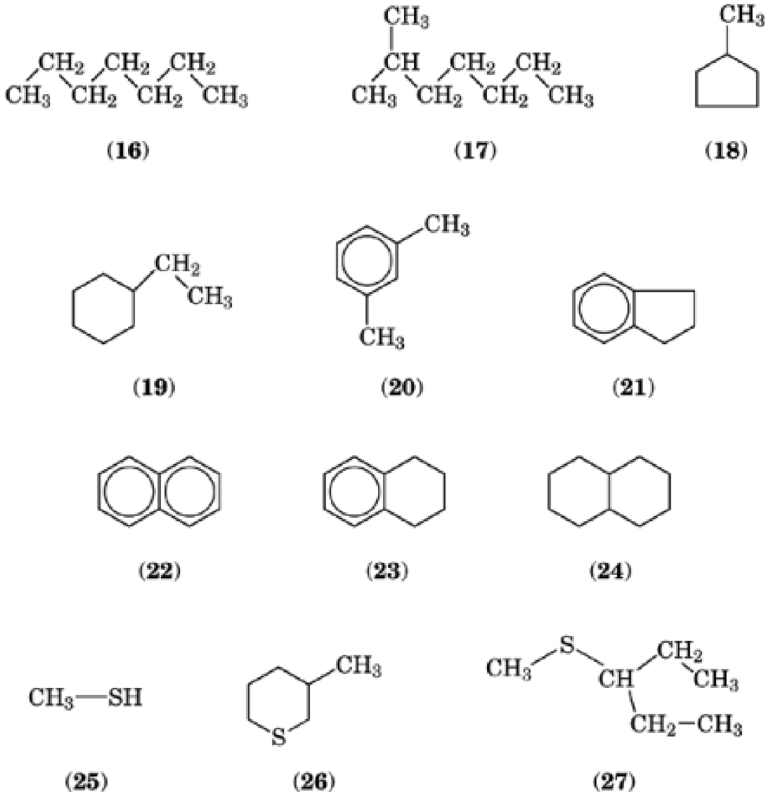


Fig. 5. Structures of compounds in petroleum naphtha. See Table 2.

has also been used for the assessment of the stage of oxidation/reduction of the environment in which ancient organisms were converted into petroleum (5).

Mono- and di-cycloparaffins having five or six carbons per ring constitute the bulk of naphthenes in the middle-distillate boiling range, decreasing in concentration as the carbon number increases (2). The alkylated naphthenes appear to have a single long side chain as well as one or more methyl or ethyl groups (27). Similarly substituted three-ring naphthenes have been detected by gc/ms (28). Generally, fused rings share just a single face; however, the highly symmetric multifused molecule adamantane and its alkyl-substituted homologues have been found (3, 29).

The most abundant aromatics in the mid-distillate are mono-, di-, and trimethyl naphthalenes. Other one- and two-ring aromatics are undoubtedly present in small quantities as either naphtheno or alkyl homologues in the C_{11}–C_{20} range. In addition to these homologues of alkylbenzenes, tetralins, and naphthalenes, the mid-distillate contains some fluorenes and phenanthrenes, with traces of biphenyls (30). The phenanthrene structure is favored over that of anthracene structure (2).

The *S*-heterocyclics in the mid-distillate range are primarily the thiacyclanes, benzothiophenes, and dibenzothiophenes. There are lesser amounts of dialkyl–, diaryl–, and aryl–alkyl sulfides (29). Alkylthiophenes are scarce or absent, but some evidence exists for benzthiacyclanes (Fig. 6). As for the naphtha fractions, these sulfur species account for a minimal fraction of the total sulfur in the crude.

Although only ppm levels of nitrogen are found in the mid-distillates, both neutral and basic nitrogen compounds have been isolated and identified in fractions boiling below 345°C (12). Pyrroles and indoles account

Table 3. Compounds Found in Petroleum Middle-Distillates[a]

Compound	CAS Registry Number	Molecular formula	Structure number
n-hexadecane (cetane)	[544-76-3]	$C_{16}H_{34}$	(**28**)
2-methylpentadecane	[1560-93-6]	$C_{16}H_{34}$	(**29**)
pristane (2,6,10,14-tetramethylhexadecane)	[1921-70-6	$C_{19}H_{40}$	(**30**)
phytane (2,6,10,14-tetramethylhexadecane)	[638-36-8]	$C_{20}H_{42}$	(**31**)
pentamethyldecalin	[80655-44-3]	$C_{15}H_{28}$	(**32**)
fichtelite	[2221-95-6]	$C_{19}H_{34}$	(**33**)
adamantane	[281-23-2]	$C_{10}H_{16}$	(**34**)
biphenyl	[92-52-4]	$C_{12}H_{10}$	(**35**)
fluorene	[86-73-7]	$C_{13}H_{10}$	(**36**)
phenanthrene	[85-01-8]	$C_{14}H_{10}$	(**37**)
2-methylbenzothiophene	[1195-14-8]	C_9H_8S	(**38**)
dibenzothiophene	[132-65-0]	$C_{12}H_8S$	(**39**)
benzothiacyclohexane	[2054-35-5]	$C_9H_{10}S$	(**40**)

[a] See 6.

for about two-thirds of the nitrogen. The remaining nitrogen is found in the basic pyridine and quinoline compounds. Most of these compounds are alkylated.

3. Vacuum Gas Oils

As is indicated in Figure 4, saturates contribute less to the vacuum gas oil (VGO) than the aromatics, but more than the polars present at percentage, rather than trace, levels. VGO itself is occasionally used as a heating oil but most commonly it is processed by catalytic cracking to produce naphtha or by extraction to yield lubricant oils.

Within the VGO saturates, distribution of paraffins, isoparaffins, and naphthenes is highly dependent on the petroleum source. The naphthenes account for roughly 60% of the saturates in a normal crude oil. However, samples can be found having paraffins from <20 to $>80\%$. In most samples, the n-paraffins from C_{20}–C_{44} are still present in sufficient quantity to be detected as distinct peaks in gc analyses. Some crude oils show a nearly symmetric pattern of peaks such that each carbon number is present in regular progression up to a maximum around C_{27}. Other crude oils show a similar distribution, but have preference for odd-numbered alkanes. Both the distribution and the selectivity toward odd-numbered hydrocarbons are considered to reflect differences in petrogenesis of the crude oils. Although n-paraffins are distinct in the gc, these usually account for only a few percent of the saturates measured by gc.

The bulk of VGO saturates consists of isoparaffins and especially naphthenes (Fig. 3). A few isoprenoid compounds, such as squalane, C_{30}; lycopane, C_{40}; and carotanes, C_{40}; have been detected. Analyses of petroleum waxes isolated from the saturates show a parallel lower level of 2- and 3-methyl alkanes as the most identifiable isoparaffins. Mass spectrometry techniques show that the naphthenes contain from one to more than six fused rings. Having an average carbon number of C_{32}, even the six-ring naphthenes have some alkyl substitution. For mono- and diaromatics, the alkyl substitution typically involves one long side chain and several short methyl and ethyl substituents. Some specific tetracyclic naphthenes, including steranes, and pentacyclic naphthenes, including hopanes, have been used as biomarkers (Fig. 7). These hopanes and steranes have also been used as nondegradable conserved internal markers for estimating biodegradation of crude oils during bioremediation processes (31).

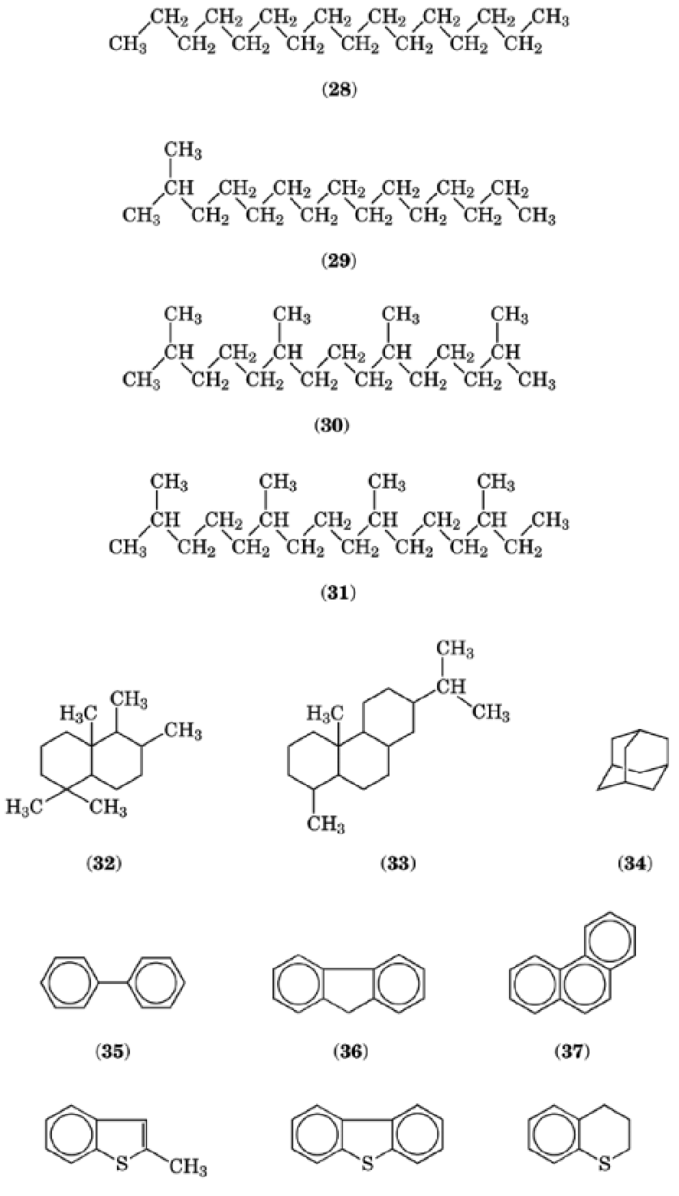


Fig. 6. Structures of compounds in petroleum middle-distillates. See Table 3.

The aromatics in VGO may contain one to six fused aromatic rings that may bear additional naphtheno rings and alkyl substituents in keeping with their boiling range. Mono- and diaromatics account for about 50% of the aromatics in petroleum VGO samples. Studies by nmr indicate an average of 3.5 substituents per monoaromatic, whereas ms techniques show the presence of up to four fused naphthenic rings on some aromatic compounds (19). This is consistent with the suggestion that these species originate from the aromatization of

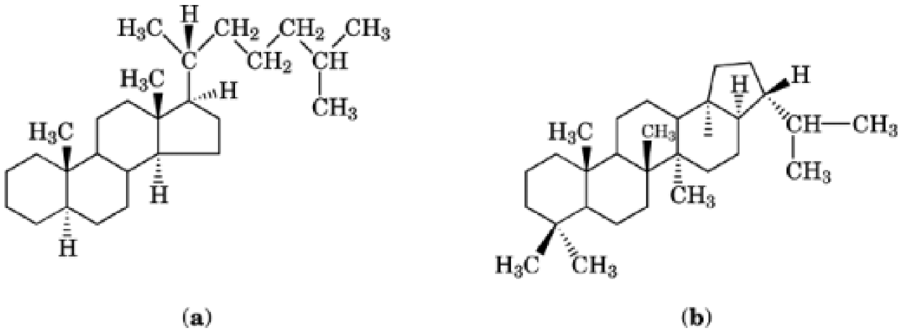


Fig. 7. Naphthenic biomarker compounds: (**a**) 5α(H),14α(H),17α(H),20(R)-cholestane [481-21-0], a C_{27} sterane, and (**b**) 17α(H),21α(H)-hopane [13849-96-2], a C_{30} pentacyclic triterpane.

biogenic steroids (2, 32). Although they are present at lower concentration, alkyl benzenes and naphthalenes commonly show one long side chain and multiple short side chains.

The fused 3_+ ring aromatics in petroleum include both *cata*- and *peri*-condensed structures (see Table 4, Fig. 8). The *cata*-condensed species are those structures where only one face is shared between rings, the *peri*-condensed molecules are those that share more than one face. The fused ring aromatics form the class of compounds known as polynuclear aromatic hydrocarbons (PAH) which includes a number of recognized carcinogens in the 4_+ ring family (33). Because of the potential health and environmental impact of PAH, these compounds have been studied extensively in petroleum.

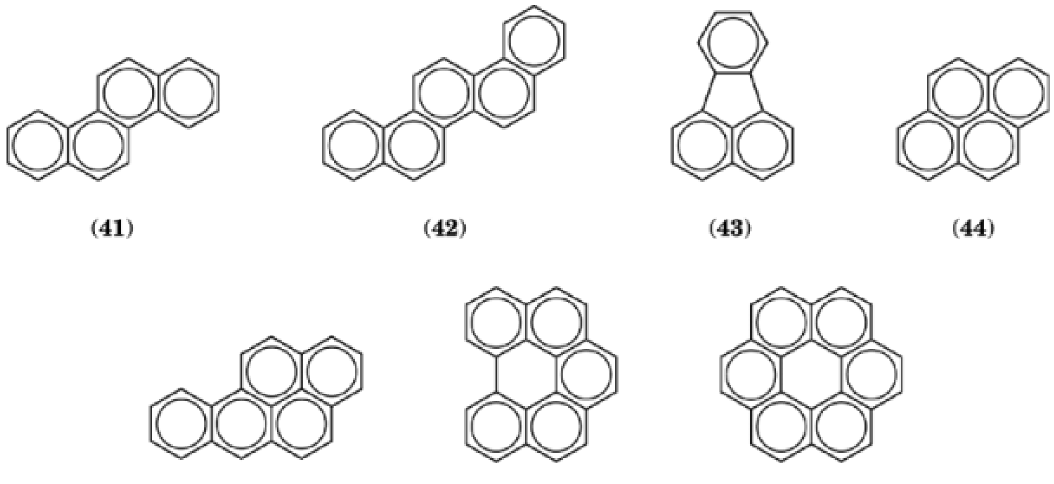
Table 4. Fused-Ring Polynuclear Aromatic Hydrocarbons Found in Petroleum[a]

Compound	CAS Registry Number	Molecular formula	Structure number
cata-Condensed aromatics			
phenanthrene	[85-01-8]	$C_{14}H_{10}$	(**37**)[b]
chrysene	[218-01-9]	$C_{18}H_{12}$	8
picene	[213-46-7]	$C_{22}H_{14}$	(**42**)
peri-Condensed aromatics			
fluoranthene	[206-44-0]	$C_{16}H_{10}$	(**43**)
pyrene	[129-00-0]	$C_{16}H_{10}$	(**44**)
benzo[α]pyrene	[50-32-8]	$C_{20}H_{12}$	(**45**)
benzoperylene	[11057-45-7]	$C_{24}H_{14}$	(**46**)
coronene	[191-07-1]	$C_{24}H_{12}$	(**47**)

[a] See Fig. 8.
[b] See Fig. 6.

The total levels of three–six ring PAH in a petroleum VGO fraction range from 2 to 25%. However, the concentrations of individual isomers that have been quantified are generally expressed in parts per million (ppm). Although the most abundant reported individual phenanthrene compounds appear to be the C_1–C_3 derivatives, the average carbon number for the phenanthrenes in a VGO is C_{32}, indicating that the average side chain actually has 18 carbon atoms. The analytical techniques that have been used to isolate individual compounds are biased toward the least substituted aromatic structures, ie, the techniques that facilitate identification of a few target compounds eliminate the majority of multiring aromatic compounds.

Within petroleum certain aromatic structures appear to be favored. For example, alkyl phenanthrenes outnumber alkyl anthracenes by as much as 100:1. In addition, despite the bias in separation methods, alkyl



(41) **(42)** **(43)** **(44)**

(45) **(46)** **(47)**

Fig. 8. Structures of fused-ring polynuclear aromatic hydrocarbons. See Table 4.

derivatives appear to be more abundant than the parent ring compounds. For larger ring PAH, patterns exist, but are more difficult to detect owing to analytical resolution limits for the increasing number of isomers. A survey of crude oils by ms suggests that chrysenes are favored over pyrenes. Even for larger ring systems, the carbon numbers show that the parent compounds are missing or present in trace quantities. The C_3 derivatives are most prevalent.

The heterocyclics are significant contributors to the VGO fraction. Assuming an average molecular weight of 320 for the VGO, sulfur levels of from 0.5 to $>2.5\%$ would indicate that from 5 to $>25\%$ of the molecules contain a sulfur atom (26). Results of API-60 show that 60% of the sulfur compounds are thiophenic and the remainder sulfidic. Although there is some contribution from alkylaromatic and diaryl sulfides, no dialkyl sulfides were detected. In contrast, from 20 to 36% of the sulfur compounds existed in thiacyclane structures of one to eight saturated rings. Gel permeation chromatography of the aromatic fractions coupled with ms led to identification of >30 homologous thiophenic series including 18 having more than a single sulfur atom. Although the distribution into these groups varied widely among crude oils, benzothiophenes, and dibenzothiophenes, having from zero to six naphtheno rings, were the prevalent thiophenic forms of sulfur, accounting for 30–40% of the total. Many of the homologous series appear to be *S*-analogues of the hydrocarbon PAH, ie, benzologues of dibenzothiophene, but no specific compounds have been isolated and identified unambiguously.

The nitrogen levels in crude oils are generally an order of magnitude lower than those of sulfur. In the VGO range, the nitrogen-containing compounds include higher molecular weight pyridines, quinolines, benzoquinolines, amides, indoles, and carbazoles; and molecules having two nitrogens, ie, diaza compounds, and three or four aromatic rings are especially prevalent (20). Typically, about one-third of the compounds are basic, ie, pyridine and its benzologues, whereas the remainder are present as neutral species such as amides and carbazoles. Although benzo- and dibenzoquinolines found in petroleum are rich in sterically hindered structures, hindered and unhindered structures have been found to be present at equivalent concentrations in source rocks. This has been rationalized as geo-chromatography in which the less polar (hindered) structures moved more readily to the reservoir (5).

Oxygen levels in the VGO parallel the nitrogen content. Thus, the most identified oxygen compounds are phenols and carboxylic acids, frequently called naphthenic acids. These may account for from ppm to nearly 3% of a VGO. The presence of numerous complex naphthenic and naphthenoaromatic acid structures in crude oils, especially immature forms, has been shown (34). Among the different structures a number of specific steroid carboxylic acids have been identified.

4. Vacuum Residua

The vacuum residua or vacuum bottoms is the most complex fraction. Vacuum residua are used as asphalt and coker feed. In the bottoms, few molecules are free of heteroatoms; molecular weights range from 400 to >2000, so high that characterization of individual species is virtually impossible. Separations by group type become blurred by the sheer mass of substitution around a core structure and by the presence of multiple functionalities in a single molecules. Simultaneously, the traditional gc and ms techniques require the very volatility that this fraction lacks.

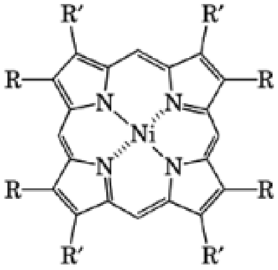
Classically, pentane or heptane precipitation is used as the initial step for the characterization of vacuum resid. The asphaltenes, the material that precipitates from the sample, range from 1 to 25% of vacuum resids. The soluble portion (maltenes) are then fractionated chromatographically into several poorly resolved classes for characterization. Asphaltene separations provide a simple way to remove some of the largest and most polar components; however, the fractions generated are still so complex that techniques such as nmr, titration, xps, or ir are mainly used to provide some average compositional detail. Asphaltenes are thought to be aggregates of complex molecular structures of MW <2000. The aggregation mechanism is not well understood. Evidence exists for both charge-transfer, among the numerous heteroatom functionalities, and φ–φ stacking of aromatics within the asphaltenes. The pitfalls of using average techniques for these complex systems have been discussed (21).

Additional general compositional information has been derived for the vacuum resid, using more sophisticated techniques that differ in approach toward isolating fraction. One approach uses functional group separations. Ion exchange (qv) has been used as the primary fractionation of vacuum resid into acids, bases, and neutrals (20, 32). A second approach uses distillation into carefully defined atmospheric equivalent boiling points (AEBP) for the initial separation (21). Fractions generated by these separations have been subjected to a battery of techniques including liquid chromatography, field ionization mass spectrometry (fims), nmr, and elemental analyses.

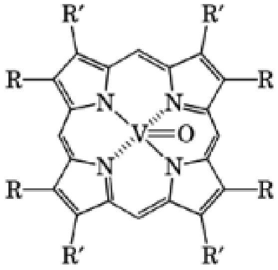
In fractions distilled above 565°C, the paraffins contribute only $<2\%$ of the 10–20% saturates. The one- to six-ring naphthenes contribute the rest, having average carbon numbers of 48–63. In the same study, aromatics shift progressively from an even distribution of mono- to tetraaromatics toward one dominated by five-ring types. As evidenced by the z-series data, many aromatics bear one or more naphtheno rings plus sufficient side-chain carbon atoms to bring the mean mol wt to nearly 800, ie, approximately 40 side-chain carbon atoms for a five-ring aromatic core of mol wt about 300. However, the elemental data show that 80% of the molecules could contain sulfur. Whereas most of the sulfur is thought to be thiophenic, studies by x-ray absorption spectroscopy have shown that in some vacuum resids it can exist in approximately equal amounts as sulfidic and thiophenic forms (35).

For the $565^{\circ}+$ fraction the levels of nitrogen and oxygen may begin to approach the concentration of sulfur. These elements consistently concentrate in the most polar fractions to the extent that every molecule contains >2 heteroatoms. At this point, structural identification is not attempted. Rather, average characterization techniques are used to confirm the presence of functionalities such as acids, phenols, carbazole, and benzoquinoline, found in lower boiling fractions. Several models have been proposed based on the observed functionalities, apparent molecular weight, and elemental analysis of the fraction. These models suggest that molecules having boiling points greater than 565°C consist of multiple units similar to lower boiling components linked together with carbon and sulfur bridges rather than ever-increasing fused-ring structures (36).

The Ni and V concentrated into the vacuum resid appear to occur in two forms. From 10 to 14% of each of these two metals can be distilled in the 565–705°C boiling range, where they exhibit the strong visible Soret bands associated with the porphyrin structure. This tetrapyrrole structure **48,49**, possibly derived from ancient chlorophyll, has been confirmed by a variety of analytical techniques.



(48)



(49)

Because the metalloporphyrins can provide insights into petroleum maturation processes, they have been studied extensively, and several families of related structures have been identified. On the other hand, the bulk of the metals are found in the heaviest fraction. Because those highly polar fractions do not exhibit a Soret band, they are designated nonporphyrin metals. However, a number of techniques suggest that the metals in these Soret-inactive compounds still exist in molecules containing tetrapyrrole structures. Whereas other structures have been proposed (37), none has been demonstrated conclusively (38).

BIBLIOGRAPHY

“Petroleum Composition,” in *ECT* 1st ed., Vol. 10, pp. 92–97, by A. N. Sachanen, Socony-Vacuum Oil Co., Inc.; in *ECT* 2nd ed., Vol. 14, pp. 845–855, by R. H. Hunt and M. J. O’Neal, Shell Oil Co.; “Petroleum–Composition,” in *ECT* 3rd ed., Vol. 17, pp. 119–131, by J. J. Elliott and M. T. Melchior, Exxon Research and Engineering Co.

Cited Publications

1. R. R. F. Kinghorn, *An Introduction to the Physics and Chemistry of Petroleum*, John Wiley & Sons, Inc., New York, 1983.
2. B. P. Tissot and D. H. Welte, *Petroleum Formation and Occurrence*, Springer-Verlag, New York, 1978.
3. J. M. Hunt, *Petroleum Geochemistry and Geology*, W. H. Freeman & Co., San Francisco, Calif., 1979.
4. J. G. Speight, *Fuel Science Technology Handbook*, Marcel Dekker, New York, 1990.
5. M. Yamamoto, *Adv. Org. Geochem.* **19**, 389 (1992).
6. K. E. Peters and J. M. Moldowan, *The Biomarker Guide: Interpreting Molecular Fossils in Petroleum and Ancient Sediments*, Prentice-Hall, Englewood Cliffs, N.J., 1993.

7. R. E. Jordan and J. R. Payne, *Fate and Weathering of Petroleum Spills in the Marine Environment*, Ann Arbor Science Publications, Ann Arbor, Mich., 1980.
8. M. F. Ali, M. Saleem, and S. M. Ghazali, *Hydrocarbon Process.* **60**(9), 129 (1981).
9. L. R. Snyder, *Accounts Chem. Res.* **3**, 290 (1970).
10. B. J. Mair, Z. Ronen, E. J. Eisenbraun and A. G. Horodysky, *Science* **154**, 1339–1341 (1966).
11. *Characterization of the Heavy Ends of Petroleum* API Project 60, American Petroleum Institute, 1971; J. E. Dooley, D. E. Hirsch, C. J. Thompson, and C. C. Ward, *Hydrocarbon Process.* **53**, 187 (1974).
12. L. R. Snyder, *Anal. Chem.* **41**, 314 (1969).
13. L. R. Snyder, B. E. Buell, and H. E. Howard, *Anal. Chem.* **40**, 1303 (1968).
14. L. R. Snyder, *Anal. Chem.* **41**, 1084 (1969).
15. D. M. Jewell and co-workers, *Anal. Chem.* **44**, 1391 (1972).
16. M. M. Boduszynski, *Energy Fuels* **2**, 597 (1988).
17. C. S. Hsu, K. Qian, and Y. C. Chen, *Anal. Chim. Acta* **264**, 79 (1992); C. S. Hsu and co-workers, *Energy Fuels* **5**, 395 (1991).
18. P. Cleon, M. C. Foucheres, D. Cagniant, D. Severin, and W. Holstein, *Chromatographia* **18**, 543 (1985); J. M. Schmitter and co-workers, *Org. Geochem.* **6**, 579 (1984).
19. K. Qian and C. S. Hsu, *Anal. Chem.* **64**, 2327 (1992).
20. J. A. Green and co-workers, *Analysis of Heavy Oils: Method Development and Application to Cerro Negro Heavy Petroleum*, NIPER-452 (DE90000200, 2 vols.), IIT Research Institute, National Institute for Petroleum and Energy Research (NIPER), Bartlesville, Okla., 1989.
21. K. H. Algelt and M. M. Boduszynski, *Composition and Analyses of Heavy Petroleum Fractions*, Marcel Dekker, New York, 1994.
22. M. M. Boduszynski, *Liquid Fuel Technol.* **2**, 395 (1984).
23. C. S. Hsu and K. Qian, *Energy Fuels* **7**, 268 (1993).
24. B. J. Mair, *Oil Gas J.*, 130 (Sept. 14, 1964).
25. H. T. Rall, C. J. Thompson, H. J. Coleman, and R. L. Hopkins, *Sulfur Compounds in Crude Oil*, Bulletin 659, U.S. Bureau of Mines, Washington, D.C., 1962.
26. V. S. Aksenov and V. F. Kamyanov, in R.Kh. Freidlina, and A. E. Skorova, eds., *Organic Sulfur Chemistry*, Pergamon Press, New York, 1980, Chapt. 1.
27. R. J. Hood, R. J. Clerc, and M. J. O'Neal, *J. Inst. Pet.* **45**, 168 (1959).
28. J. S. Richardson and D. Miller, *Anal. Chem.* **54**, 765 (1982).
29. St. Lanka and Sl. Hala, *Erdoel Kohle Erdgas Petrochem.* **11**, 698 (1958).
30. S. W. Lee, S. Coulombe, and B. Glavincevski, *Energy Fuels* **4**, 20 (1990).
31. R. C. Prince and co-workers, *Environ. Sci. Technol.* **28**, 142 (1994); C. S. Hsu, unpublished data.
32. J. F. McKay, J. H. Weber, and D. R. Latham, *Anal. Chem.* **48**, 891 (1976).
33. A. Dipple, in C. E. Searle, ed., *Chemical Carcinogens*, ACS Monograph 173, American Chemical Society, Washington, D.C., 1976, Chapt. 5, 245–314.

34. W. K. Seifert and R. M. Teeter, *Anal. Chem.* **42**, 750 (1970).
35. G. N. George and M. L. Gorbaty, *J. Am. Chem. Soc.* **111**, 3182 (1989).
36. I. A. Wiehe, *Energy Fuels* **8**, 536 (1994).
37. R. H. Fish and J. Komlenic, *Anal. Chem.* **56**, 510 (1984).
38. C. D. Pearson and J. B. Green, *Energy Fuels* **7**, 338 (1993).

WINSTON K. ROBBINS
Exxon Research and Engineering Company
CHANG SAMUEL HSU
Exxon Research and Engineering Company

Related Articles

Nomenclature In the Petroleum Industry; Coal; Gas, natural; Tar and pitch