Selecting a reduced suite of diagnostic ratios calculated between petroleum biomarkers and polycyclic aromatic hydrocarbons to characterize a set of crude oils

R. Fernández-Varela, J.M. Andrade*, S. Muniategui, D. Prada

Dept. Analytical Chemistry, University of A Coruña, Campus da Zapateira s/n, E-15071 A Coruña, Spain

ARTICLE INFO

ABSTRACT

A set of 34 crude oils was analysed by GC–MS (SIM mode) and a suite of 28 diagnostic ratios (DR) calculated. They involved 18 ratios between biomarker molecules (hopanes, steranes, diasteranes and triaromatic steroids) and 10 quotients between polycyclic aromatic hydrocarbons. Three unsupervised pattern recognition techniques (i.e., principal components analysis, heatmap hierarchical cluster analysis and Kohonen neural networks) were employed to evaluate the final dataset and, thus, ascertain whether the crude oils grouped as a function of their geographical origin. In addition, an objective variable selection procedure based on Procrustes Rotation was undertaken to select a reduced set of DR that comprised for most of the information in the original data without loosing relevant information. A reduced set of four DR (namely; TA21, D2/P2, D3/P3 and B(a)F/4-Mpy) demonstrated to be sufficient to characterize the crude oils and the groups they formed.

1. Introduction

As industrialization progresses worldwide, more petroleum resources are required. Since demand increases and prices raise and they depend partly on the origin of the oil, there is a need for reliable analytical methodologies to screen whether a given oil batch corresponds to the contract agreement. It follows that chemical characterization of raw crude oils is a hot ongoing topic because of the strong efforts petrochemical companies undergone to exploit current productive extraction fields to their utmost. Disappointingly, chemical characterization is quite complex because the crude oil extracted at different wells in a same production field may be different. Even within a well the crude oil collected at different depths may be different because of the different mixture ratios than can occur within the source rocks [1].

The so-called petroleum biomarkers (or biological markers) are particularly useful to objectively assess the provenance of an oil [2] and their analysis was recommended to derive important information on the crude oils (oil fingerprinting), to differentiate oils, to search for the source of a spillage and to monitor the degradation process and the weathering stage of an oil under a variety of conditions [3]. Biomarkers are naturally occurring, ubiquitous and stable hydrocarbons that appear in crude oils and most petroleum products [2]. They derive from formerly living organisms whose organic materials were preserved in oil source rocks that upon burial (heat and pressure) generated crude oil over geologic time.

In this work hopanes, steranes, diasteranes and triaromatic steroids were analysed and they will be introduced briefly. In general, terpanes and steranes are branched cycloalkanes consisting of multiple condensed five- or six-carbon rings [4]. Hopanes are pentacyclic triterpanes which contain 27–35 carbon atoms. They derive from bacterial (prokaryotic) membrane lipid precursors. There, cyclization of squalene precursors give rise to hopanoids as, for instance, bacteriohopanetetrol [5]. Most hopanes derive from the most abundant one, C_{35} tetrahydroxybacteriohopane [6]. Steranes, formally named perhydrocyclopentanophenanthrene rings, are a class of 4-cyclic compounds derived from steroids or sterols (which constitute half of the lipids in the lipid membranes in all eukaryotic cells) via diagenetic and catagenetic degradation and saturation. The relative abundances of C_{27}-, C_{28}- and C_{29}-steranes in oils reflect the carbon number distribution of the sterols in the organic matter in the source rocks [2]. Diasteranes are rearranged steranes that have no biological precursors, and are most likely formed during diagenesis and catagenesis. Triaromatic steroids can originate by aromatization and loss of a methyl group from monoaromatic steroids which, in turn, derived exclusively from sterols with a side-chain double bond during early diagenesis [5].

* Corresponding author. Tel.: +34 981 167000; fax: +34 981 167065.
  E-mail address: andrade@udc.es (J.M. Andrade).

Despite biomarkers can be used as such it was found more useful to calculate ratios among them [2,7,8], which were termed diagnostic ratios (DR). Besides, diagnostic ratios are unaffected by short-term weathering processes as long as they are based on compounds with little or comparable susceptibility to weathering. Some of them are known to relate to the thermal maturity of the source rocks that gave rise to a crude oil; the type of organic matter that was present in the source rock (e.g., terrestrial vs. marine); and/or the degree of weathering that may have occurred in the crude oil reservoir before oil extraction. Their specificity, diversity, complexity, and relative resistance to weathering make them useful 'markers' in the characterization of spilled oils, candidate source oils, and background contamination [9]. Understanding the meaning of these ratios, and evaluating whether they can be of use to characterize different oil production areas is therefore a must for petrochemical companies and researchers on environmental characterization of spilled hydrocarbons [2,9–11].

The diagnostic ratios considered in this study were based upon '*genetically*' significant (source-specific) variables that are known to occur among crude oils from different geologic basins as well as on several previous studies from different research groups [1,2,7–10,12–16].

In addition, the large amount of data generated in many environmental studies and/or petroleum quests require the use of multivariate chemometric tools to provide a unbiased, objective and defensible means to differentiate among qualitatively similar oils [9,10,17,18]. Chemometric approaches are also required to extract a reduced set of DRs which may suffice to still differentiate among the different types of oils.

In this paper 40 biomarkers and 16 polycyclic aromatic hydrocarbons (PAHs) were analysed in 34 crude oils by gas chromatography with mass spectrometry detection (GC–MS), using the SIM mode. In total, 28 DRs were calculated and studied by unsupervised pattern recognition techniques (principal components analysis, heatmap hierchical cluster analysis and Kohonen neural networks) in order to assess whether they may differentiate different petroleum-producing geographical areas worldwide. Further, an objective variable selection technique based on Procrustes Rotation was applied to extract a minimum set of relevant DRs to differentiate among the different petroleum basins as much as possible. Procrustes Rotation was selected because of its ability to select analytical variables instead of abstract factors or combinations of variables like other chemometric techniques do.

2. Experimental

2.1. Samples

Thirty-four crude oil samples representing different petroleum bearing basins throughout the world were collected from several Spanish refineries (Table 1). Their specific gravities ranged from 19° to 48° API. Crude oil samples were water and sediment extracted following an ASTM guide [19], light-protected and stored at 4°C until analysis. The analytes were measured after dissolving 20–50 mg of each sample, weighted accurately in an analytical balance, in 5 mL of dichloromethane (Super purity solvent, Merck).

2.2. Gas chromatography–mass spectrometry

An HP 6890 instrument (Agilent Technologies, Palo Alto, CA, USA) with a pulsed splitless injector, an HP 5973 mass spectrometry detector and an HP-5MS fused silica capillary column (J&W Scientific, Folsom, CA, USA) 60 m long (0.25 mm i.d., 0.25 µm film thickness) were employed. Operating conditions were: starting oven temperature, 40°C, held isothermally for 1 min, and raised

Table 1
Resume of the crude oils employed in this study.

Origin		Product
North Africa	Libya	Amna, Es Sharara, Sarir, Sirtica
	Algeria	Sahara Blend
	Tunisia	Ashtart
Middle East	Azerbaijan	Azeri Light
	Iran	Foroozan, Soroosh
	Saudi Arabia	Arabian Heavy
	Syria	Syria
Central Africa	Nigeria	Brass, Ea, Escravos
	Ecuatorial Guinea	Zafiro
North Sea	North Sea	Brent, Draugen, Ekofisk, Flotta, Forties, Gullfaks, Norne, Schiehallion, Statfjord
South America	Argentina	Cañadón Seco
	Colombia	Caño Limón, Vasconia
	Ecuador	Oriente
	Venezuela	Santa Barbara
South Africa	Angola	Girassol
Central America	México	Maya
Russia	Russia	Siberian Light, Tengiz, Ural Light

to 300°C at 6°C/min and held isothermally for 30 min. Carrier gas: Helium, 1 mL/min constant flow. Injector and transfer line temperature were 300 and 280°C, respectively. Ionization energy: 70 eV, ion source temperature, 230°C. Injection was performed in the pulsed splitless mode, injected sample: 1 µL. The *m/z* range for MS analysis was 40–440. The SIM mode (selected ion monitoring) was used throughout. In total, 20 hopanes, 13 steranes and diasteranes, 7 triaromatic steroids biomarkers and 16 PAHs were analysed (see Table 2 for more details).

As mentioned in the introduction, different diagnostic ratios (DRs) have been proposed in literature to differentiate crude oils and the most common ones were selected to perform this work: 27Ts, 28ab, 25nor30ab, 29Ts, 30O, 30G, 29ab, 30d, 32abS, 27dia, 29aaS, 29bb, 27bbSTER, 28bbSTER, 29bbSTER, TA21, TA26, TA27, D2/P2, D3/P3, D3/C3 and Retene/P4 [20]. Besides, common diagnostic PAHs include dibenzothiophenes and phenanthrenes, although some other possibilities exist [13,21,22]. Thus, 'source-specific' marker compounds, including alkylated PAH hydrocarbons within homologous alkylation isomeric groups were identified as well and their ratios calculated. 'Source-specific' here means that the DRs may serve as unambiguous markers for some oils under study, as many times they are subject to little interference from absolute concentration fluctuation of individual compounds [15]. The 'source-specific' DRs considered here were 2-MP/1-MP and 4-MD/1-MD [20], B(a)F/4-Mpy, B(b+c)F/4-Mpy, 2-Mpy/4-Mpy and 1-Mpy/4-Mpy [23].

Further, in order to calculate the DRs a previous internal quality control evaluation was done, as it had been shown that biomarkers may be affected by the analytical variability and sample heterogeneity [8,20]. All samples were analysed by triplicate and the relative standard deviation (RSD) of each compound calculated. Then, following [20] and [23], only DRs for which the RSDs of the compounds involved were lower than 5% were employed. Accordingly, a suite of 28 DR (18 quotients between the peak heights of several biomarkers and 10 ratios between peak areas for several PAHs) were calculated. Their full description is displayed in Table 3.

2.3. Chemometric techniques and software

Here unsupervised pattern recognition multivariate techniques had to be used because the lack of more samples of known origin impeded us to get independent validation sets of samples that might be used to fully validate supervised methods. Hence, three unsupervised methods were selected. Two of them, principal com-