Journal of Chromatography A, 1380 (2015) 162–170

Contents lists available at ScienceDirect

Journal of Chromatography A

journal homepage: www.elsevier.com/locate/chroma

Screening of oil sources by using comprehensive two-dimensional gas chromatography/time-of-flight mass spectrometry and multivariate statistical analysis

Wanfeng Zhang[a,b], Shukui Zhu[a,b,*], Sheng He[b], Yanxin Wang[a,**]

[a] State Key Laboratory of Biogeology and Environmental Geology, China University of Geosciences, Wuhan 430074, China
[b] Key Laboratory of Tectonics and Petroleum Resources of Ministry of Education, China University of Geosciences, Wuhan 430074, China

ARTICLE INFO

Article history:
Received 24 July 2014
Received in revised form
19 December 2014
Accepted 22 December 2014
Available online 2 January 2015

Keywords:
GC × GC/TOFMS
Principal component analysis
Cluster analysis
Oil screening
Petroleum

ABSTRACT

Using comprehensive two-dimensional gas chromatography coupled to time-of-flight mass spectrometry (GC × GC/TOFMS), volatile and semi-volatile organic compounds in crude oil samples from different reservoirs or regions were analyzed for the development of a molecular fingerprint database. Based on the GC × GC/TOFMS fingerprints of crude oils, principal component analysis (PCA) and cluster analysis were used to distinguish the oil sources and find biomarkers. As a supervised technique, the geological characteristics of crude oils, including thermal maturity, sedimentary environment etc., are assigned to the principal components. The results show that tri-aromatic steroid (TAS) series are the suitable marker compounds in crude oils for the oil screening, and the relative abundances of individual TAS compounds have excellent correlation with oil sources. In order to correct the effects of some other external factors except oil sources, the variables were defined as the content ratio of some target compounds and 13 parameters were proposed for the screening of oil sources. With the developed model, the crude oils were easily discriminated, and the result is in good agreement with the practical geological setting.

1. Introduction

It is well-known that screening of oil sources play important roles for petroleum exploration. In order to conceive the processes of oil migration and to understand the origin, thermal maturity, biodegradation level, as well as oil–oil and oil-source rock correlations, a conclusive characterization of crude oil samples is the first and foremost task. As we know, crude oils contain a limited number of classes of compounds, but the number of individual components of which these fractions consist is enormous. In addition, the composition of crude oils is susceptible to physical (e.g. evaporation, emulsification, natural dispersion, dissolution and sorption), chemical (photo degradation) and biological (mainly microbial degradation) weathering processes [1]. As one of the most complex fluids, both advanced separation tools and high efficient methods of data processing are essential in order to achieve accurate molecular fingerprints of crude oils with different origin or geochemical history and find potential marker compounds.

Currently, the widely used separation tools for the determination of components in crude oils are gas chromatography coupled to mass spectrometry (GC/MS) or tandem mass spectrometry (GC/MS/MS) [2–4]. As described in many earlier literatures, the peak capacity of conventional one-dimensional gas chromatography (1DGC) is by far insufficient to reveal the full complexity of crude oils and so peaks overlap seriously, even with the use of highly selective stationary phases and high resolution 'narrow-bore' columns. Therefore, some labor-intensive and time-consuming sample preparation steps have to be used before GC analysis. Even so, it has been concluded that no single chromatographic technique is able to separate or characterize these complex mixtures completely and a multidimensional separation system has to be considered [5].

Comprehensive two-dimensional gas chromatography (GC × GC) is one of the multidimensional separation methods. In contrast to 1DGC, GC × GC provides enhanced peak capacity and resolving power, with the use of two capillary columns of different stationary phases serially connected by a modulator. It has been regarded as an ideal technique for the analysis of complex mixtures where compounds of similar chemical structure

* Corresponding author at: 388 Lumo Road, China University of Geosciences, Wuhan 430074, China. Tel.: +86 27 67883067.
** Corresponding author at: 388 Lumo Road, China University of Geosciences, Wuhan 430074, China. Tel.: +86 27 67883998.
E-mail addresses: shukuizhu@126.com (S. Zhu), yx.wang@cug.edu.cn (Y. Wang).

http://dx.doi.org/10.1016/j.chroma.2014.12.068
0021-9673/

are grouped into distinct patterns in the GC × GC plane providing useful information on both their boiling point and polarity [6]. Especially when coupled with time-of-flight mass spectrometry (TOFMS), the fast data acquisition rate, full-range mass sensitivity and deconvolution of overlapped peaks become available, which makes the identification of compounds more credible [7]. The above advantages of GC × GC/TOFMS are very suitable to the separation of complex mixtures like crude oils.

However, to distinguish the oil sources and find potential biomarkers, a series of crude oil samples have to be analyzed and then many complex GC × GC/TOFMS chromatograms containing tens of thousands of peaks are yielded. It is definitely difficult to find marker compounds from a large quantity of compounds which exactly reflect the differences among oil samples. Many existing methods usually discriminate crude oil samples or assess the thermal history, depositional environment, and the type of organic matter by comparing peak areas of two or more chemical compounds. The limitation to this approach is the excessive reliance on a relatively small number of biomarkers to characterize such complex fluids as crude oils [8]. Therefore, high efficient methods of data processing, such as multivariate statistical analysis, can be of great use for the enormous data sets obtained from GC × GC/TOFMS. As we know, multivariate statistical analysis can provide better options to achieve valuable information. By reducing the number of dimensions while retaining as much as possible of the data's variation, we can focus on some components containing the majority of the data's variation, instead of investigating thousands of original variables. So far, there have been some successful applications on processing GC × GC data with multivariate statistical analysis. Zhang et al. [9] identified 450 compounds in different tea samples by using GC × GC/TOFMS, and discovered the key components for distinguishing the three types of tea with significant difference by partial least squares (PLS). Purcaro et al. [10] applied an iterative approach to fully exploit the informative content of a GC × GC/MS data set. The most informative compounds were identified and collected in a "blueprint" of specific defects and a powerful discrimination of samples was obtained in view of a sensory quality assessment. Rocha et al. [7] applied GC × GC/TOFMS combined with principal component analysis (PCA) and partial least squares discriminant analysis (PLS-DA) to the untargeted and comprehensive study of the volatile composition of human urine, and the results suggest that the urinary volatile profiles may be useful for differentiating subjects with different physiological conditions. Schmarr et al. [11] developed an image processing approach for the processing of GC × GC data, and the achieved profiles have been used for multivariate statistical analysis and allowed clustering of comparable sample origins and prediction of unknown samples.

In this study, a screening method of oil sources has been developed based on GC × GC/TOFMS analysis and multivariate statistical analysis. To build an oil screening model, some parts of oil samples are obtained from different reservoirs in the same oil field and the other samples from different oil fields, and the origin and geochemical history are known in advance. Based on the molecular fingerprints of oil samples, principal component analysis (PCA) and cluster analysis were used to distinguish the oil sources and find potential biomarkers with combination of supervised techniques.

2. Experimental

2.1. Samples and chemicals

A set of 15 crude oil samples were obtained from different reservoirs in Jianghan basin and Nanyan basin, China. Silica gel (100–200 mesh, activated at 150 °C for 8 h before use) and alumina (50–100 mesh, activated at 400 °C for 4 h before use) were purchased from Shanghai Nahui Reagent Co., Shanghai, China. Dichloromethane, *n*-hexane, ethanol and chloroform were of chromatographic grade quality and purchased from Tedia Co., Fairfield, OH, USA. Medical absorbent cotton was rinsed by chloroform until no fluorescence. Two custom standard mixtures, hydrocarbon window defining standard (0.2 mg/mL for each) and polycyclic aromatic hydrocarbon (PAH) solution mixture (0.2 mg/mL for each), were purchased from AccuStandard (New Haven, PA, USA). Hydrocarbon window defining standard includes *n*-C_8-*n*-C_{40} alkanes, Pristane, and Phytane. PAH solution mixture includes 15 components, i.e., naphthalene ($C_{10}H_8$), acenaphthene ($C_{12}H_{10}$), fluorene ($C_{13}H_{10}$), phenanthrene ($C_{14}H_{10}$), anthracene ($C_{14}H_{10}$), fluoranthene ($C_{16}H_{10}$), pyrene ($C_{16}H_{10}$), benzo[a]anthracene ($C_{18}H_{12}$), Chrysene ($C_{18}H_{12}$), benzo[b]fluoranthene ($C_{20}H_{12}$), benzo[k]fluoranthene ($C_{20}H_{12}$), benzo[a]pyrene ($C_{20}H_{12}$), indeno[1,2,3-cd]pyrene ($C_{22}H_{12}$), dibenz[a,h]anthracene ($C_{22}H_{14}$), and benzo[g,h,i]perylene ($C_{22}H_{12}$).

2.2. Sample preparation

In GC × GC/TOFMS analysis, 30 mg of crude oil sample was dissolved in 50 mL of *n*-hexane, and then dispersed by ultrasonic treatment for 5 min. After put for 12 h at room temperature, the solution was transferred to a funnel for filtering asphaltenes. The asphaltenes on the absorbent cotton were further rinsed by *n*-hexane for three times. All the filtrates were collected and carefully concentrated under nitrogen flow to 0.5 mL for analysis.

In GC/MS analysis, a micro chromatographic column (30 cm × 10 mm I.D.) was prepared using mixed stationary phase of activated silica gel and alumina at a ratio of 3:2 by referring to relative literature [2]. It was pre-conditioned with 20 mL of *n*-hexane before use. The crude oil samples were firstly pretreated using the same steps as GC × GC/TOFMS analysis, i.e. 30 mg of crude oil sample was dissolved in *n*-hexane, transferred to a funnel for filtering asphaltenes, and concentrated to 0.5 mL. Then the concentrated sample was transferred to the top of the micro chromatographic column for further separation. The saturated hydrocarbon fraction was eluted with *n*-hexane (25 mL), aromatic fraction with a mixture of *n*-hexane: dichloromethane (1:2, v/v, 25 mL) and polar fraction including some heteroatomic compounds with a mixture of ethanol:chloroform (1:1, v/v, 20 mL), respectively. The fractions were carefully concentrated under nitrogen flow to 0.5 mL for analysis.

2.3. Instrumentation and methods

The GC × GC system consisted of a GC (7890A model, Agilent Technologies, Wilmington, DE, USA) equipped with a secondary oven, a quad-jet dual stage modulator. The GC oven contained two capillary columns that were connected serially by means of a Siltek treated universal press-tight connector (Restek Corp., Bellefonte, PA, USA). Nitrogen and air were used as the cold and hot gases, respectively. A time-of-flight mass spectrometer (Pegasus 4D, Leco Corp., St. Joseph, MI, USA) was used to acquire mass spectral data of the effluents from the GC × GC. A DB-5MS column 50 m × 0.25 mm × 0.25 μm and a DB-17Ht column 1.8 m × 0.10 mm × 0.10 μm (J&W Scientific, Folsom, CA, USA) were used as the primary and secondary dimensional columns, respectively. High purity helium (99.9995%) was used as carrier gas at a flow rate of 1.0 mL/min. The inlet temperature was 300 °C. Injections were performed in the splitless mode, and the injection volume was 0.2 μL. All injections were done with an Agilent 7683B autosampler. The 1st oven temperature was initially held at 50 °C for 1 min, ramped to 100 °C at 10 °C/min, and then to 300 °C (30 min hold) at 2 °C/min. The 2nd oven temperature was 10 °C