A Novel Method for Chemical Fingerprinting of Oil and Petroleum Products Based on Electrospray Mass Spectrometry and Chemometrics

Ingvar Eide* and Kolbjørn Zahlsen

Statoil Research Centre, N-7005 Trondheim, Norway

A novel method for the chemical fingerprinting of oil and petroleum products has been developed. The method is based on full-scan mass spectrometry, using a single quadrupole liquid chromatography–mass spectroscopy (LC-MS) instrument, but with direct injection (no chromatographic separation) and without fragmentation of the molecules. In the present work, positive electrospray ionization has been used (ESI-MS). Pattern recognition of the spectra is performed using multivariate data analysis (chemometrics). The mass spectrometer is operated in the scan mode, in the mass number (m/z) range of 65–1000, and gives spectra with one distinct line per integer mass number. One average spectrum is obtained from each analysis. Oils, even the most complex heavy crude oils, can be analyzed directly without pretreatment, except dissolution in dichloromethane. One analysis requires 1 min to perform. Principal component analysis is used to evaluate similarities and differences between oil samples. The projections to latent structures method is used to correlate chemical fingerprints with observed properties. The method has successfully been applied to distinguish between different crude oils and mixtures of crude oils, and it seems to be very promising for a variety of other applications.

Introduction

Chemical fingerprinting of oils and petroleum products implies the provision of a preferentially unique compositional pattern that can be recognized, primarily by multivariate pattern recognition techniques (chemometrics).

Many fingerprinting techniques are based on the identification and quantification of a limited number of a priori selected compounds and is, for example, used for oil spill identification.[1] Other techniques such as infrared (IR) and near infrared (NIR) spectroscopy give spectra for crude oils and petroleum products;[2] however, the spectra cannot be resolved into individual mass numbers or compounds. Fingerprinting, in terms of a more complete characterization of complex mixtures of organic compounds has been used on extracts of diesel exhaust particles.[3] The study applied full-scan gas chromatography–mass spectrometry (GC-MS), followed by curve resolution, to obtain peaks and spectra for individual compounds. However, chromatographic separation and curve deconvolution become very complicated with oils and petroleum products, which are extremely complex mixtures. Furthermore, a drawback with GC is that large molecules (with boiling points (bp) of >400–450 °C) are difficult to analyze, and the meth-

odology is time- and resource-demanding. Electrospray ionization mass spectrometry (ESI-MS) is an alternative. Direct infusion into a quadrupole mass spectrometer has been used to obtain the spectra of fossil fuels[4,5] and detergents in gasoline.[6] Direct infusion into a high-resolution mass spectrometer has been used to obtain spectra of heterocyclic organic compounds in crude oils and coal, with the main goal to elucidate chemical structure.[7–9] Multivariate pattern recognition was not used in these studies.

The aim of the present study was to develop a quick and simple method for the chemical fingerprinting of oil and petroleum products, based on the combined platform of single quadrupole ESI-MS methodology (direct injection) and chemometrics. The methodology has been used to discriminate between different crude oils and mixtures of oils. Principal component analysis (PCA) was used to evaluate similarities in composition between mixtures.[10] The projections to latent structures (PLS)[11] method was used for the regression modeling,

* Author to whom correspondence should be addressed. Telephone: +47 90997296. Fax: +47 73967286. E-mail: ieide@statoil.com.

(1) Daling, P. S.; Faksness, L.-G.; Hansen, A. B.; Stout, S. A. *Environ. Forensics* **2002**, *3*, 263–278.
(2) Aske, N.; Kallevik, H.; Sjöblom, J. *Energy Fuels* **2001**, *15*, 1304–1312.
(3) Eide, I.; Neverdal, G.; Thorvaldsen, B.; Grung, B.; Kvalheim, O. *Environ. Health Perspect.* **2002**, *110* (Suppl. 6), 985–988.

(4) Zahn, D.; Fenn, J. B. *Int. J. Mass Spectrom.* **2000**, *194*, 197–208.
(5) Qian, K.; Edwards, K. E.; Diehl, J. H.; Green, L. A. *Energy Fuels* **2004**, *18*, 1784–1789.
(6) Carraze, B.; Delafoy, J.; Bertin, J.; Beziau, J.-F.; Lange, C. M. *Rapid Commun. Mass Spectrom.* **2004**, *18*, 451–457.
(7) Hughey, C. A.; Rodgers, R. P.; Marshall, A. G. *Anal. Chem.* **2002**, *74*, 4145–4149.
(8) Wu, Z.; Jernström, S.; Hughey, C. A.; Rodgers, R. P.; Marshall, A. G. *Energy Fuels* **2003**, *17*, 946–953.
(9) Marshall, A. G.; Rodgers, R. P. *Acc. Chem. Res.* **2004**, *37*, 53–59.
(10) Jackson, J. E. *A User's Guide to Principal Components*; Wiley: New York, 1991.
(11) Wold, S.; Ruhe, A.; Wold, H.; Dunn, W. J., III. *SIAM J. Sci. Stat. Comput.* **1984**, *5*, 735–743.
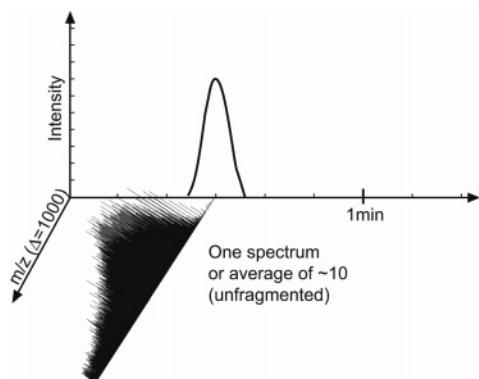
Figure 1. Illustration of the full-scan mass spectrometry without chromatographic separation.

because it overcomes the problems of intercorrelated predictor variables and data matrices where the number of variables exceeds the number of samples.[12,13]

Experimental Methods

Samples of four different crude oils (labeled A, B, C, and D) were used in the present study. Oils A and B were from the Norwegian sector in the North Sea, oil C was from South America, and oil D was from central Asia. In one experiment, the aim was to investigate the ability of the method to discriminate between the four different crude oils, and each of the four oils was analyzed separately. In another experiment, the aim was to investigate the ability of the method to express differences between binary mixtures of the two Norwegian oils quantitatively. Therefore, five different mixtures of oils A and B were analyzed (0%, 25%, 50%, 75%, and 100% of each). Oil A was a light, low-sulfur crude oil with API = 36.8°. Oil B was a medium-density, low-sulfur crude oil with API = 31.9°.

The samples were dissolved in dichloromethane (DCM) (2 mg/mL) and analyzed by full-scan mass spectrometry on an Agilent 1100 Series LC/MSD system (Agilent Technologies Inc., Palo Alto, CA). The system consisted of a model G1322A mobile-phase degassing unit, a model G1311A quaternary pump with a gradient mixer for up to four mobile-phase constituents, a model G1367A autosampler, and a model G1946D single quadrupole mass spectrometer. Samples with a volume of 1 μL were injected by the autosampler and led into the mass spectrometer by a 70-cm length of PEEK tubing (inner diameter (I.D.) of 0.18 mm), without separation on a chromatographic column. Each sample was injected (analyzed) 10 times. The mass spectrometer was operated in the scan mode in the mass number (m/z) range of 65–1000, using atmospheric-pressure electrospray positive ionization (AP-ESI). The mobile phase consisted of acetonitrile and ammonium acetate (50 mM) in a ratio of 90:10, and the mobile-phase flow rate was 0.2 mL/min. The fragmentor voltage was 100 V.

With direct injection, each analysis requires 1 min and gives only one peak in the chromatographic direction (Figure 1). One average spectrum was obtained from each individual analysis, calculated from ∼10 individual spectra obtained at half peak height, after background subtraction. The process of background subtraction and acquisition of spectra was performed by a post-run macro to ensure identical data collection between different injections. The mass spectrometric analysis was performed to avoid fragmentation. The spectra consequently reflect those molecules in the samples that have been ionized. Each average spectrum was tabulated to a row with numbers,
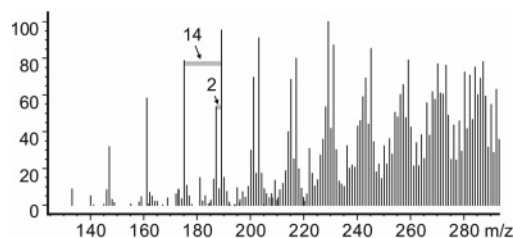
(12) Kvalheim, O. M. *Anal. Chim. Acta* **1989**, *223*, 53–73.
(13) Kettaneh-Wold, N. *Chemom. Intell. Lab. Syst.* **1992**, *14*, 57–69.

Figure 2. Expanded image of a portion of the mass spectrum depicted in Figure 1.

where each number represents the height of each spectral line. The unit resolution of the ESI-MS system gives m/z values that are accurate to one decimal place. To construct a compressed matrix from all the individual analysis, the mass numbers are rounded off to integer mass numbers.

In the present work, each of the four crude oils were analyzed 10 times, and, consequently, the final matrix X contains 40 rows (one row per analysis) and 931 columns (one per integer mass number with abundance values above zero). The five combinations of oils A and B resulted in a matrix X with 50 rows and 934 columns, and a corresponding matrix Y with 50 rows and two columns (where the percentage of oils A and B are given in each column).

Prior to multivariate analysis, the abundance data (spectral line values) were normalized to a constant sum within each row and the normalized data were centered. Multivariate data analysis was performed with Simca-P 10.5 (Umetrics, Umeå, Sweden). PCA was performed on matrix X for the evaluation of similarities between mixtures.[10] Regression modeling was performed with the PLS method to correlate chemical fingerprints (matrix X) to the percentages of oils A and B in the binary mixtures (matrix Y).[11] The PCA and PLS models were validated with respect to explained variance and goodness of prediction (shown as Q^2), and the latter was obtained after cross-validation.[14] The PLS model was also evaluated with respect to goodness of fit (R^2).

Results and Discussion

Figure 1 illustrates the full-scan mass spectrometry without chromatographic separation. Although the spectrum may appear to be unresolved, there is one distinct line per integer mass number (m/z), as can be observed when focusing on a smaller part of the spectrum (Figure 2). Each line represents a limited number of compounds with the same mass number. Figure 2 illustrates the detailed, fine structure that can be systematically explored by multivariate data analysis. Repetitive spacings of 14 Da (CH_2) and 2 Da (saturated versus double-bond analogues) are indicated in the figure.

Figure 3 shows spectra from the four crude oils A, B, C, and D. Although the overall profiles of the spectra are similar, differences can be observed. Figure 4 shows the score plot obtained after PCA of the fingerprint data from 10 analyses of each of the four crude oils. The first principal component explains 53% of the variation in the data, and the second component explains 23%. (Another 6% can be explained with three components and $Q^2 = 0.75$.) The score plot illustrates the very high repeatability of the fingerprint data, and that the four oils are significantly different.

The PLS analysis of the fingerprint data of the mixtures of oils A and B versus the blend matrix resulted in a very good PLS model with a goodness of fit of $R^2 = 0.996$ and a goodness of prediction of $Q^2 = $

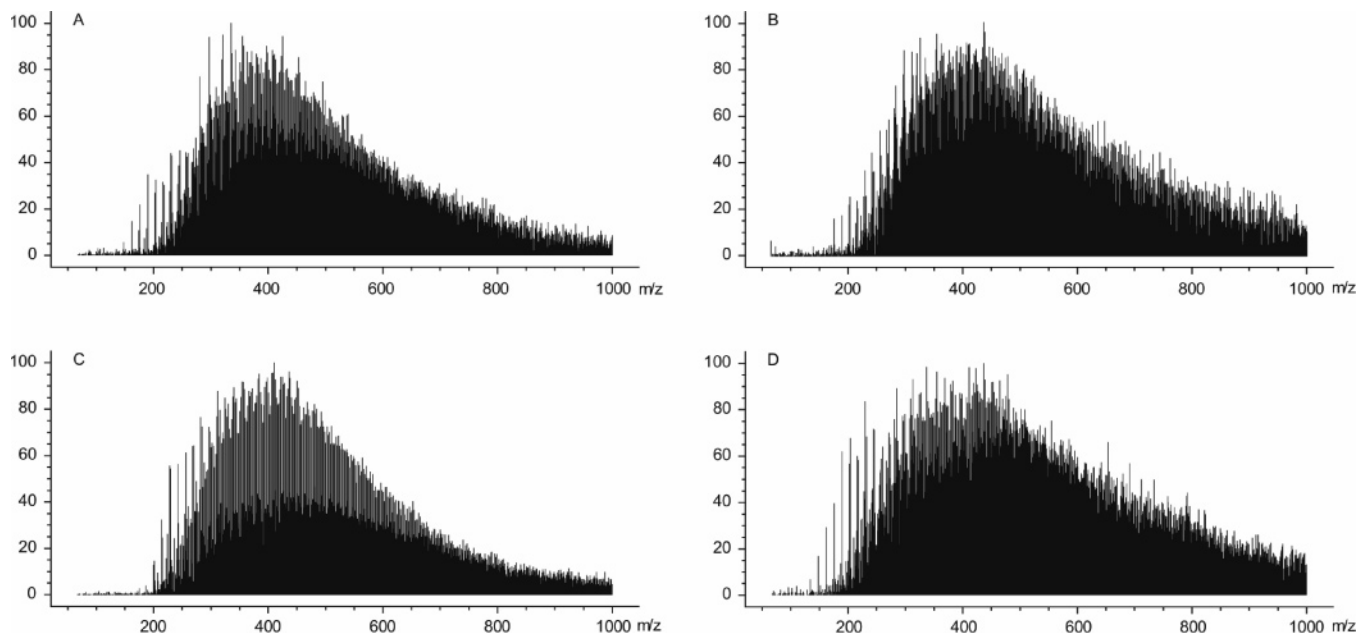(14) Wold, S. *Technometrics* **1978**, *20*, 397–405.

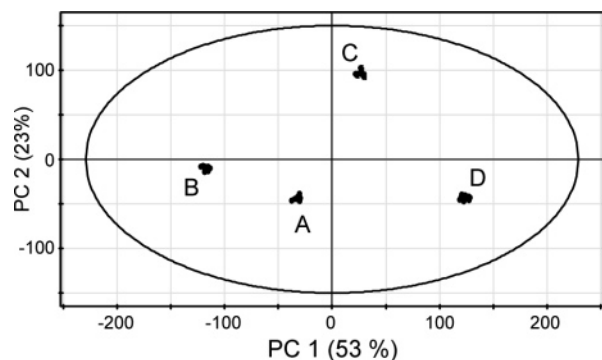**Figure 3.** Spectra of the four different crude oils (A, B, C, and D).

Figure 4. Score plot obtained after fingerprinting and multivariate pattern recognition of the four crude oils (each sample was analyzed 10 times).
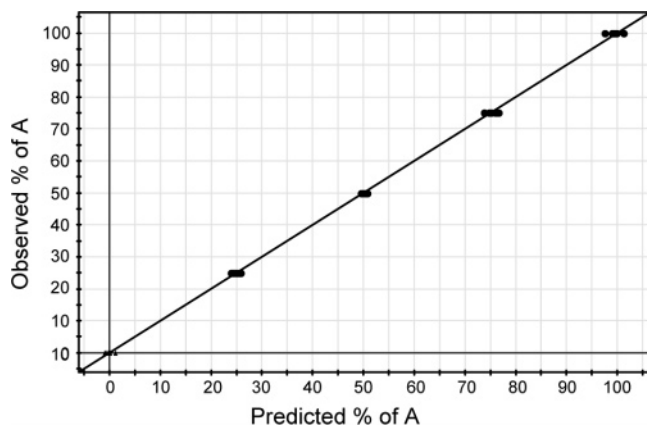

Figure 5. Observed versus predicted percentage of oil A in a binary mixture of oils A and B (each mixture was analyzed 10 times).

0.985 with two PLS components. Figure 5 shows the observed versus predicted percentage of oil A in the binary mixture and illustrates the excellent calibration curve.

These two examples demonstrate that the mass spectra obtained from the single quadrupole liquid chromatography−mass spectroscopy (LC-MS) instrument contains sufficient information to discriminate between oils and mixtures of oils and correlate the spectral patterns to the percentage of each oil in the binary mixtures. The spectra represent those parent compounds (not fragments) that have been ionized with the actual ionization technique. In the present work, atmospheric-pressure electrospray positive ionization was used. Other ionization techniques may be used to detect other groups of compounds (complementary fingerprinting).

High-resolution mass spectrometry (Fourier transform ion cyclotron resonance−mass spectrometry) has recently been applied to the analysis of crude oils and coal with the main goal to elucidate the chemical structure.[7−9,15] The high-resolution methodology is unique and provides valuable background information for our fingerprinting technique. The compounds detected by positive electrospray ionization are primarily nitrogen-, oxygen-, or sulfur-containing heterocyclic organic compounds. This has recently been shown also with single quadrupole ESI-MS, using model substances.[5] Each line in our spectra may represent up to 30−40 different compounds or groups of compounds. High-resolution mass spectrometry may produce ∼10 times as many lines in the spectra, compared to the single quadrupole technique. Hence, each line in the high-resolution spectra may still represent many individual compounds with the same mass number. Furthermore, because of the fact that commonly encountered elemental compositions only yield certain masses, the high-resolution spectra seem to be dominated, to a large extent, by one large line per mass number, smaller neighboring lines, and gaps.[7,9,15] As a consequence, a major part of the spectral patterns will be captured by the single quadrupole technique, although the smaller lines, in reality, are added into one large line per mass number. The single quadrupole spectra have enough complexity and repeatability and contain sufficient information for the pattern recognition technique. However, high-resolution mass spectrometry is superior, in regard to structure

(15) Qian, K.; Rodgers, R. P.; Hendrickson, C. L.; Emmett, M. R.; Marshall, A. G. *Energy Fuels* **2001**, *15*, 492−498.

elucidation. It should be emphasized that, with our method, we in fact are rounding off the decimals from tabulated mass spectra to integer mass numbers. This gives condensed data matrixes with few empty cells, which is advantageous for the multivariate pattern recognition.

The high-resolution mass spectrometry technique demonstrates and confirms several specific spectral characteristics of complex hydrocarbon-containing mixtures, such as the repetitive spacing of 14 Da (CH_2) and 2 Da (saturated versus double-bond analogues), which are also observed at unit resolution with our methodology (see Figure 2). The high-resolution mass spectrometry technique also states that electrospray ionization produces singly charged ions.[8,9] The profile of our spectra, the similarity with the condensed high-resolution spectra, and the assumption that the compounds are singly charged implies that most of the compositional information is contained in the *m/z* range of 65−1000. However, the single quadrupole mass spectrometer can be operated in the scan mode in the *m/z* range of 65−3000, in intervals of more than 1000. The single quadrupole ESI-MS technology is very suitable for the purpose of chemical fingerprinting: The method captures most of the information needed for pattern recognition. Single quadrupole instruments are commercially available and the number is rapidly increasing worldwide. The instruments are reliable and cost-effective. There are challenges ahead to improve methodology, verify reproducibility, standardize analytical conditions, improve data preprocessing, etc. We have demonstrated that, with the combination of ESI-MS and chemometrics, within a series, we can distinguish between oils and mixtures of oils with impressive repeatability. We have not yet reached a performance analogous to that we have with GC-MS, where the provision of searchable spectra of fragmented individual molecules are obtained by highly standardized GC-MS. The ESI-MS spectra represent unfragmented molecules from very complex mixtures of oil, and we are presently studying the effects of a variety of parameters on the ESI-MS spectra, as well as standardization and control of these parameters. Furthermore, we are presently developing the methodology further and are evaluating a variety of different applications, including the following:

(1) Characterization of crude oils, including heavy crudes and petroleum products

(2) Estimation of the contribution of oil from various reservoirs in commingled production scenarios

(3) Detection of the effects of well treatment chemicals on crudes

(4) Source identification (e.g., of oil spills) by comparing samples of spilled oil and candidate oils

(5) Identification of compounds that make samples different (e.g., contaminants and additives)

(6) Time-based studies (hydrocarbon upgrading, degradation, oil-in-water, oil weathering, etc.) and process control

(7) Correlation of chemical fingerprints with measured physicochemical or environmental oil properties

Conclusion

The single quadrupole electrospray ionization−mass spectroscopy (ESI-MS) technology is very suitable for the purpose of chemical fingerprinting: The method captures most of the information needed for pattern recognition. Single quadrupole instruments are commercially available, and the number is rapidly increasing worldwide. The analyses are quick and easy to perform, because of direct injection, and the use of autosampler injection allows high-throughput analysis of multiple samples. The method is not limited by complexity or molecular size. Oils, including heavy crude oils, can be analyzed directly without pretreatment, except dissolution in an organic solvent (dichloromethane). The method provides fingerprints with very high repeatability. These fingerprints are characteristic for the different oils and mixtures of oils, and the patterns can be recognized by multivariate data analysis, with a powerful ability to discriminate between different oil samples. Furthermore, the chemical fingerprints can be correlated to measured properties.