

Data Interpretation Methods for Petroleomics

Annana Islam^a, Yunju Cho^a, Arif Ahmed^a, and Sunghwan Kim^{a,b*}

^aKyungpook National University, Department of Chemistry, Daegu, 702-701, Korea

^bGreen-Nano Materials Research Center, Daegu, 702-701, Korea

Received September 4, 2012; Revised September 12, 2012; Accepted September 12, 2012

First published on the web September 20, 2012; DOI: 10.5478/MSL.2012.3.3.63

I. Introduction	63
II. Kendrick mass defect plot	64
III. Van Krevelen diagram	64
IV. Double-bond equivalence vs. carbon number plot	65
V. Application of statistical analysis	66
VI. Correlation analysis	66
VII. Conclusions and future studies	67
VIII. Acknowledgement	67
IX. References	67

Abstract: The need of heavy and unconventional crude oil as an energy source is increasing day by day, so does the importance of petroleomics: the pursuit of detailed knowledge of heavy crude oil. Crude oil needs techniques with ultra-high resolving capabilities to resolve its complex characteristics. Therefore, ultra-high resolution mass spectrometry represented by Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS) has been successfully applied to the study of heavy and unconventional crude oils. The analysis of crude oil with high resolution mass spectrometry (FT-ICR MS) has pushed analysis to the limits of instrumental and methodological capabilities. Each high-resolution mass spectrum of crude oil may routinely contain over 50,000 peaks. To visualize and effectively study the large amount of data sets is not trivial. Therefore, data processing and visualization methods such as Kendrick mass defect and van Krevelen analyses and statistical analyses have played an important role. In this regard, it will not be an overstatement to say that the success of FT-ICR MS to the study of crude oil has been critically dependent on data processing methods. Therefore, this review offers introduction to petroleomic data interpretation methods.

Keywords: Mass spectrometry, Laser desorption, Petroleum, Data interpretation

Introduction

Energy has been playing a key role in the development of modern human society for many years. A person in modern industrialized society consumes approximately ten times more energy than one in the agricultural society because modern human is relying on energy for food production, heating, construction, and transportation.¹⁻³ Fossil fuels especially crude oil have been one of the most heavily used energy resources. It is well-known that oil is a limited resource and hence people are continuously looking for new alternative energy.^{4,5} However, the transition of our society to adapt to new energy source will take decades⁶ and it is logical to identify and use more immediately usable energy resources. In fact, as the world's crude oil deposit becomes heavier, it is less efficient to utilize the

crude oils.⁷ It is because smaller amount of economically viable component is generated from heavy crude oils than lighter ones. Therefore, it is very important to devise methods to increase the efficiency of utilizing heavy crude oils.⁸

Understanding the heavy component of crude oil at the molecular level has been very important to improve petroleum processing.^{9,10} Petroleomics refers to a research effort where detailed knowledge of heavy crude oil is pursued.^{11,12} High resolution mass spectrometry especially Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS) has been one of the key components for petroleomics.¹³ It has not been easy to study the heavy (typically molecular weight over 400 Da) and/or polar compounds by use of traditional analytical methods such as gas chromatography mass spectrometry (GC-MS).¹⁴ However, Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS) coupled to various ionization sources has enabled us to observe and study heavy

*Reprint requests to Dr. Prof. Sunghwan Kim
E-mail: sunghwank@knu.ac.kr

compounds up to about 800 Da with and without polar functional groups.^{15–17} For example, atmospheric pressure photo ionization (APPI) coupled to FT-ICR MS is a powerful tool to study aromatic and/or sulfur containing compounds.^{10,18} More than thousands of compounds have been routinely observed by use of this technique.¹⁹

The successful application of FT-ICR MS to study crude oil at a molecular level has been possible partly due to key advances in data processing methods.^{20–22} If the data processing methods were not been developed, it would have been very laborious to process and study thousands of peaks contained in each crude oils' spectra. The complex high-resolution mass spectra are the results of the complex nature of petroleum. It may happen that, spectrum of over 100,000 peaks comes out from a single crude oil sample. Therefore, development of data interpretation methods has played a crucial role in the study of crude oils by high-resolution MS.²³ For an example, Kendrick mass defect (KMD) and van Krevelen analytical methods have been important data processing methods to simplify and visualize crude oils' spectra.^{24,25} There are other important methods developed to visualize the complex spectra as well.^{26,27} Therefore, the objective of this paper is to provide a review of developments related to data interpretation of very complex spectra of crude oil provided by FT-ICR MS.

Kendrick mass defect plot

Calculation of Kendrick mass is done by multiplying the observed m/z values by ratio of nominal mass/exact mass of a given functional group.²⁸ Typically, CH_2 is used as a functional group for the KMD calculation. For CH_2 KMD, 14.0000/14.01565 is multiplied to the observed m/z values. Nominal Kendrick mass is calculated by rounding up or down Kendrick mass. The Kendrick mass defect (KMD) is calculated by subtracting the Kendrick mass from the nominal Kendrick mass. Therefore, the digits after the decimal point of the Kendrick mass define the KMD values. In case of CH_2 KMD, adding or subtracting $(\text{CH}_2)_n$ from a given molecular formula does not change KMD values of the formulae. In other words, Kendrick mass values of elemental formulae differing only by $(\text{CH}_2)_n$ will differ from each other only by whole numbers. In summary, each series of peaks differing only by $(\text{CH}_2)_n$ will have the same and their own unique KMD value. For examples, the KMD values of benzene (C_6H_6), toluene (C_7H_8) and phenol ($\text{C}_6\text{H}_6\text{O}$) can be calculated as follows. In the examples given below, benzene and toluene have the same KMD value because their elemental composition is different by CH_2 but the phenol has the different KMD value.

$$\text{KMD}(\text{C}_6\text{H}_6) = 78.04695 \times (14.0000/14.01565) - 78 = 0.9598 \quad (1)$$

$$\text{KMD}(\text{C}_7\text{H}_8) = 92.06300 \times (14.0000/14.01565) - 92 = 0.9598 \quad (2)$$

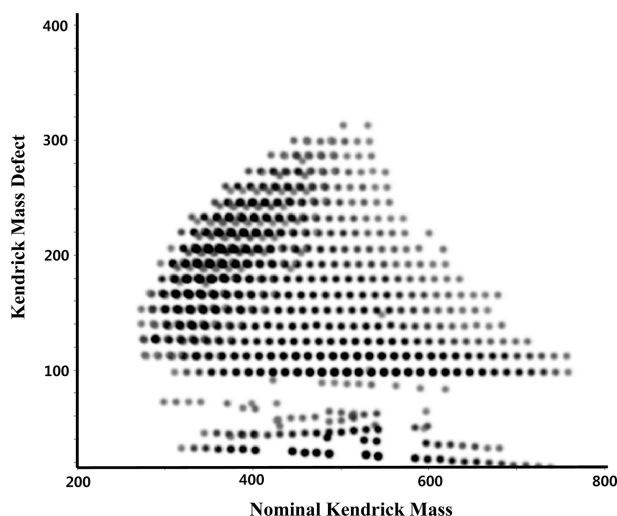


Figure 1. Kendrick mass defect diagram of a crude oil.

$$\text{KMD}(\text{C}_6\text{H}_6\text{O}) = 94.0418 \times (14.0000/14.01565) - 94 = 0.9368 \quad (3)$$

KMD plot is typically generated by plotting nominal KMD vs KMD values. An example of a KMD plot is presented in Figure 1. The dots aligned in each line parallel to x-axis have elemental compositions differing from each other by $(\text{CH}_2)_n$. The KMD values are particularly useful because they can be calculated directly from mass numbers even before the elemental compositions are calculated and assigned.²⁵ This feature allows ones to use KMD values to sort the mass numbers of calibrated mass spectra with a given neutral units. Relative abundance of peaks can be presented by color or size of dots.

A method analyzing KMD in higher-order was developed and reported.²⁹ For the analysis, mass numbers from spectra are firstly grouped by KMD values (e.g., by the CH_2 series), and then the groups are further sorted by the second KMD series (e.g., by the H_2 series). A group of peaks in different classes are plotted with the CH_2 -based KMD values as the abscissa and the ratio of the CH_2 -based KMD and H_2 -based KMD of the CH_2 -based KMD as the ordinate.

Van Krevelen diagram

The van Krevelen diagram was originally used to study bulk elemental analysis data of coals.³⁰ In the original van Krevelen plot, bulk hydrogen-to-carbon (H/C ratio) ratio was plotted as the ordinate and the bulk oxygen-to-carbon ratio (O/C ratio) as the abscissa. In this way, each sample was plotted as a dot in the diagram.³⁰ Later, Kim et al. applied the van Krevelen diagram to plot the elemental compositions obtained by FT-ICR MS.²⁴ To construct a van Krevelen diagram out of information obtained from high resolution mass spectrometry, accurate mass numbers obtained from the spectra are first converted into elemental formulae. Secondly, each of the molar ratio of hydrogen-to-

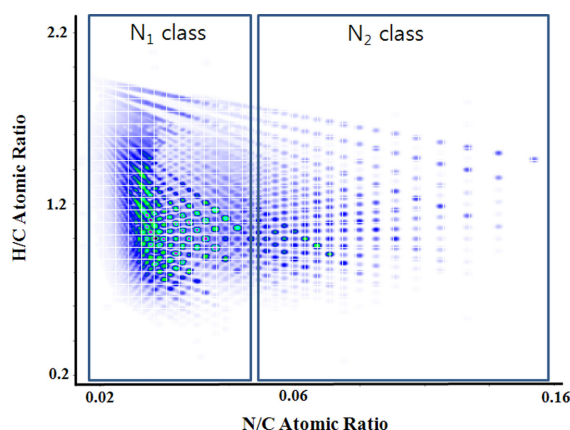


Figure 2. Van Krevelen diagram of nitrogen containing (N_1 and N_2) classes.

carbon (H/C ratio) and the molar oxygen-to-carbon ratio (O/C ratio) of the formulae are plotted as the ordinate and the abscissa. In this way of plotting, each peak (or formulae) observed in a crude oil spectrum is plotted as a dot in the diagram. Van Krevelen diagram can be used to plot heteroatom classes and each heteroatom class can be plotted by these Krevelen diagrams. The van Krevelen diagram can be used to estimate major components observed in complex mass spectra.²⁴

The relative abundance of peak observed in a given spectrum can be color-coded and presented as a contour plot.³¹ An example of van Krevelen diagram constructed from a crude oil spectrum is presented in Figure 2. In the diagram, molar hydrogen-to-carbon ratios (H/C ratio) were plotted as the ordinate and molar nitrogen-to-carbon ratios (N/C ratio) were plotted as the abscissa. The van Krevelen diagram is very effective in displaying classes containing the same hetero atoms but with different numbers of heteroatoms. The N_1 and N_2 classes of compounds are displayed in the diagram shown in Figure 2 and it is clear that they are separated in the diagram by N/C ratio.

To plot formulae in the van Krevelen plot, the elemental formulae must have the same types of hetero atoms. Therefore, the van Krevelen plot can't be used to compare different heteroatom classes. For an example, O_2 and N_1 classes can't be compared by use of this technique.

Double-bond equivalence vs. carbon number plot

Double-bond equivalence (DBE) represents the number of double bonds and rings in a given molecular formula and can be calculated by the following equation for elemental formulae $C_cH_hN_nO_oS_s$:

$$DBE = c - h/2 + n/2 + 1 \quad (4)$$

DBE is very important because it enables us to predict chemical structures from elemental formulae. For example, a compound with benzene ring structure has DBE value

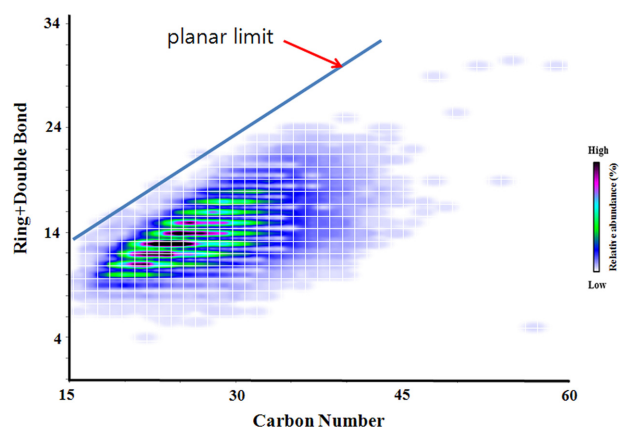


Figure 3. DBE vs carbon number plot and the planar limit observed in the plot.

of 4. Compounds each with naphthalene and anthracene core structure will have DBE value of 7 and 10. This means that addition of aromatic ring would increase DBE value by 3. Therefore, if there are series of elemental formulae which are different from each other by DBE value of 3, one can predict that the series of compounds can have aromatic structures.

The DBE values calculated from elemental formulae can be plotted against carbon number. The plot is often called "DBE vs carbon number plot". In the plot, the relative abundance of each peak can be color coded. An example of DBE vs carbon number plot generated from high resolution mass spectrum of a crude oil is shown in Figure 3. DBE vs carbon number plot can be a useful tool to figure out structures of compounds existing in crude oil sample.³² Especially, the concept of planar limits can be used for structural interpretation. In a given DBE vs carbon number plot, planar limits can be defined as a line connecting maximum observed DBE values with carbon numbers.^{33,34} The planar limit is marked in Figure 3. The structural features of the observed peaks are responsible for the variation of the slopes and intercepts of the planar limits. Structures of molecules existing in the saturates, aromatics, resins and asphaltene (SARA) fractions were proposed based on the slopes and intercepts planar limits and observed elemental formulae.³²

Another interesting concept originated from DBE vs carbon number plot is the compositional boundary.^{35,36} The compositional boundary indicates the maximum DBE values that any synthetic or natural chemical compounds can have. The line defining the compositional boundary can be calculated by the following equation.^{35,36}

$$\text{The compositional boundary} = \text{carbon number} + 1 \quad (5)$$

The compositional boundary can be used when elemental formulae are assigned. It was reported that 10% of the possible elemental formulae for masses below 1000 Da

could be excluded after the concept of compositional boundary was applied.^{35,36}

Application of statistical analysis

In petroleomic study, it is very important to analyze multiple samples and compare the obtained high resolution mass spectra. It is because the relationship between chemical and/or physical properties of crude oils and mass spectral information cannot be fully understood only with a few samples. Instead, many crude oil spectra must be analyzed and interpreted. This means that a large number of peaks easily exceeding 1,000,000 peaks have to be processed at a time. Therefore, it is indeed a great analytical challenge to extract relational information between observed peaks and the properties of crude oils. Statistical analyses have been successfully used to better understand large data sets, and hence it is reasonable to expect that statistical analyses can be successfully applied to study large amount of petroleomic data.

A statistical analysis program was developed and applied it to study 20 different samples.³⁷ In the previous study, principal component analysis (PCA) was successfully applied to group the samples based on their chemical compositions and enabled the identification of compositional differences. Additionally, hierarchical analysis (HCA) was successfully used to compare the samples. Figure 4 shows the results obtained from HCA of petroleum samples. The resulting data are presented by heat map and clustering.

For the statistical analyses to be more effective, data obtained by petroleomic technique should be more quantitative. However, at this moment, data provided by FT-ICR MS is semi-quantitative at best. Therefore, there

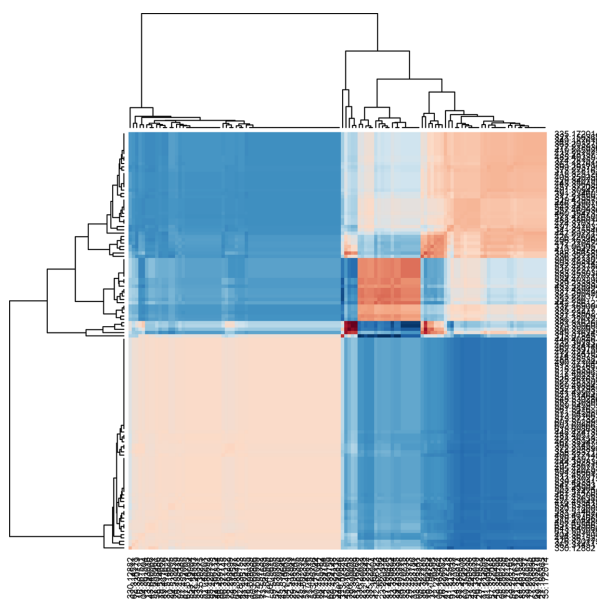


Figure 4. Diagram showing heat map and clustering resulted from hierarchical clustering analysis (HCA) of crude oil spectra.

should be more effort devoted to improve quantitative nature of petroleomic data.

Correlation analysis

Correlation is one of the important statistical methods by which relationship between two variables are identified. The correlation analysis has been applied to verify the validity of the key assumption of petroleomics that the spectral findings from high resolution mass spectrometry are related to crude oil properties.³⁸ If the assumption is not valid, the practical usefulness of petroleomics will be greatly limited. Therefore, the assumption is very important for petroleomics. The assumption was validated by seeking correlational relationships between peaks identified by high-resolution MS and the chemical/physical properties of crude oils.³⁸

The result of correlation analysis was presented using a Circos diagram (Figure 5). The Circos diagram was originally developed for genomic research and it is a very effective in visualizing complex data. In the diagram shown in Figure 5, the outside shell designates heteroatom classes. Dots located in the second shell inside the class shell represent peaks in the heteroatom classes with significant correlation with the property. The peaks located near the line just inside the second shell have correlation value (P value) of +1 and the ones located near the third shell have correlation value of -1. The peaks with P value of +1 mean that they have positive correlation with the property and ones with -1 denote negative correlation. The circle at the center shows the distribution peaks in the studied spectra.

In the previous study,³⁸ it was shown that high resolution mass spectra of crude oils showed correlational relationships with important chemical and physical properties of crude

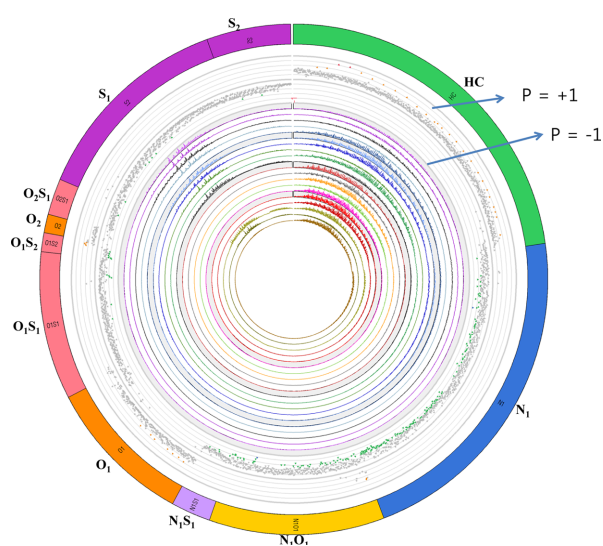


Figure 5. Circos diagram showing correlational relationship between high resolution mass spectral peak information and physical property of crude oil.

oils such as sulfur and nitrogen contents, and total acid number. Therefore, this opens up the door for chemical-component-based prediction of the properties of crude oils.

Conclusions and future studies

The application of FT-ICR MS to the analysis of crude oils has begun a new era to improve the knowledge of these materials at the molecular-level. However, these complex mixtures still remain mysterious with many issues. For example, quantitative understanding of the numerous compounds observed by FT-ICR MS is still very difficult. Future research will need to be focused on (i) quantitative interpretation, (ii) improving separation, and (iii) combining data obtained with other techniques such as ion mobility mass spectrometry to do structural interpretation.^{39,40}

Acknowledgement

This work was supported by project no. PM56951 (grants-in-aid from the Ministry of Land, Transport and Maritime Affairs, ROK). The authors thank the Korea Basic Science Institute for instrument time of 15T FT-ICR MS instrument

References

- Schaefer, C.; Weber, C.; Voss, A. *Energy* **2003**, 28, 411.
- Klass, D. L. *Energy Policy* **2003**, 31, 353.
- Thoma, M. *Energy Econ.* **2004**, 26, 463.
- Dahlquist, E.; Thorin, E.; Yan, J. *Int. J. of Energy Res.* **2007**, 31, 1226.
- Szklo, A.; Schaeffer, R. *Energy* **2006**, 31, 2513.
- Barrow, M. P. *Biofuels* **2010**, 1, 651.
- Headley, J. V.; Peru, K. M.; Barrow, M. P. *Mass Spectrom. Rev.* **2009**, 28, 121.
- Ortiz-Cruz, A.; Rodriguez, E.; Ibarra-Valdez, C.; Alvarez-Ramirez, J. *Energy Policy* **2012**, 41, 365.
- Hsieh, M.; Philp, R. P.; del Rio, J. C. *Org. Geochem.* **2000**, 31, 1581.
- Chiaberge, S.; Fiorani, T.; Savoini, A.; Bionda, A.; Ramello, S.; Pastori, M.; Cesti, P. *Fuel Proc. Tech.*
- Hsu, C. S.; Hendrickson, C. L.; Rodgers, R. P.; McKenna, A. M.; Marshall, A. G. *J. Mass Spec.* **2011**, 46, 337.
- Rodgers, R. P.; McKenna, A. M. *Anal. Chem.* **2011**, 83, 4665.
- Marshall, A. G.; Rodgers, R. P. *Proc. Natl. Acad. Sci.* **2008**, 105, 18090.
- Yassaa, N.; Meklati, B. Y.; Brancaleoni, E.; Frattoni, M.; Cicciooli, P. *Atm. Environ.* **2001**, 35, 787.
- Kujawinski, E. B. *Environ. Foren.* **2002**, 3, 207.
- Miyabayashi, K.; Naito, Y.; Tsujimoto, K.; Miyake, M. *Int. J. Mass Spec.* **2004**, 235, 49.
- Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S. *Mass Spec. Rev.* **1998**, 17, 1.
- Niedner-Schatteburg, G.; Šilha, J.; Schindler, T.; Bondybey, V. E. *Chem. Phys. Lett.* **1991**, 187, 60.
- Marshall, A. G.; Rodgers, R. P. *Acc. Chem. Res.* **2004**, 37, 53.
- Xian, F.; Corilo, Y. E.; Hendrickson, C. L.; Marshall, A. G. *Int. J. Mass Spec.* **2012**, 325–327, 67.
- Carlsohn, E.; Ångström, J.; Emmett, M. R.; Marshall, A. G.; Nilsson, C. L. *Int. J. Mass Spec.* **2004**, 234, 137.
- Rodgers, R. P.; Hendrickson, C. L.; Emmett, M. R.; Marshall, A. G.; Greaney, M.; Qian, K. *Canadian J. of Chem.* **2001**, 79, 546.
- Blakney, G. T.; Hendrickson, C. L.; Marshall, A. G. *Int. J. Mass Spec.* **2011**, 306, 246.
- Kim, S.; Kramer, R. W.; Hatcher, P. G. *Anal. Chem.* **2003**, 75, 5336.
- Hughey, C. A.; Hendrickson, C. L.; Rodgers, R. P.; Marshall, A. G. *Anal. Chem.* **2001**, 73, 4676.
- Gorshkov, M. V.; Nikolaev, E. N. *Int. J. Mass Spec. Ion Proc.* **1993**, 125, 1.
- Kazazic, S.; Zhang, H.-M.; Schaub, T. M.; Emmett, M. R.; Hendrickson, C. L.; Blakney, G. T.; Marshall, A. G. *J. Am. Soc. Mass Spec.* **2010**, 21, 550.
- Kendrick, E. *Anal. Chem.* **1963**, 35, 2146.
- Roach, P. J.; Laskin, J.; Laskin, A. *Anal. Chem.* **2011**, 83, 4924.
- van Krevelen, D. *Fuel* **1950**, 269.
- Bae, E.; Na, J. G.; Chung, S. H.; Kim, H. S.; Kim, S. *Energy Fuels* **2010**, 24, 2563.
- Kim, Y. H.; Kim, S. *J. Am. Soc. Mass Spec.* **2010**, 21, 386.
- Cho, Y.; Kim, Y. H.; Kim, S. *Anal. Chem.* **2011**, 83, 6068.
- Purcell, J. M.; Merdrignac, I.; Rodgers, R. P.; Marshall, A. G.; Gauthier, T.; Guibard, I. *Energy Fuels* **2010**, 24, 2257.
- Hsu, C. S.; Lobodin, V. V.; Rodgers, R. P.; McKenna, A. M.; Marshall, A. G. *Energy Fuels* **2011**, 25, 2174.
- Lobodin, V. V.; Marshall, A. G.; Hsu, C. S. *Anal. Chem.* **2012**, 84, 3410.
- Hur, M.; Yeo, I.; Park, E.; Kim, Y. H.; Yoo, J.; Kim, E.; No, M. H.; Koh, J.; Kim, S. *Anal. Chem.* **2010**, 82, 211.
- Hur, M.; Yeo, I.; Kim, E.; No, M. H.; Koh, J.; Cho, Y. J.; Lee, J. W.; Kim, S. *Energy Fuels* **2010**, 24, 5524.
- Fernandez-Lima, F. A.; Becker, C.; McKenna, A. M.; Rodgers, R. P.; Marshall, A. G.; Russell, D. H. *Anal. Chem.* **2009**, 81, 9941.
- Ahmed, A.; Cho, Y. J.; No, M.-H.; Koh, J.; Tomczyk, N.; Giles, K.; Yoo, J. S.; Kim, S. *Anal. Chem.* **2010**, 83, 77.