Compositional Analysis of Naphtha by FT-Raman Spectroscopy

Min-Sik Ku and Hoeil Chung*

NIR Project TFT, Production Technology Center, SK Corporation, 110 Nam-Gu, Kosa-Dong, Ulsan 680-130, Korea
Received October 22, 1998

Three different chemical compositions of total paraffin, total naphthene, total aromatic content in naphtha have been successfully analyzed using FT-Raman spectroscopy. Partial least squares (PLS) regression has been utilized to develop calibration models for each composition from Raman spectral bands. The PLS calibration results showed good correlation with those of gas chromatography (GC). Using PLS regression, the spectral information related to each composition has been successfully extracted from highly overlapped Raman spectra of naphtha.

Introduction

Raman spectroscopy has been emerging as a feasible quantitative analytical method recently, especially in refinery and petrochemical fields.[1~3] Previously it has been mostly confined to qualitative analyses such as structural identification, while quantitative analyses have only been performed using a simple band intensity or area ratio. This is partly because Raman scattering intensity is rather weak and the reproducibility of Raman spectrum is degraded due to the variation of excitation laser intensity or sample matrix changes. However, due to improvements in laser technology and other instrumental improvements, Raman spectra are now well-defined and quantitative analysis is possible for complex mixture samples, especially when combined with multivariate regression methods.[4] Especially FT (Fourier Transform) Raman spectroscopy,[5] which utilizes a near-infrared (NIR) laser as an excitation source, can be employed for an on-line process analysis because it can be easily interfaced with optical fiber for remote monitoring by guiding excitation and scattered radiation.[6,7] Additionally the fluorescence problem with visible lasers can be greatly diminished by using a low energy (longer wavelength) excitation source. Several investigations on FT-Raman spectroscopy for the analysis of products in the refinery and petrochemical areas has been reported.[8,9]

The major goal of this study is to evaluate FT-Raman spectroscopy combined with partial least squares (PLS) regression[4] for the analysis of total p̲araffin, n̲aphthene, and a̲romatic content (PNA analysis) in naphtha. Since naphtha is a very complex mixture from hundreds of individual components of C5 to C9 paraffin, naphthene, and aromatic hydrocarbons, ordinary band intensity ratio or simple regression methods are not adequate to build calibration models. Multivariate calibration techniques such as PLS regression are required to resolve this mixture and quantify desired concentrations. PLS regression is the method of reconstructing the spectral data to give the greatest variation related to analyte. It involves projecting spectral data onto a new coordinate defined by factors (eigenvectors). The set of variables (scores) in the new coordinate system is then used for calibration. Factors in PLS regression are being built using both spectral data and concentration information to give the greatest spectral variation and best correlation with concentration; therefore, PLS can extract the analyte-related information and generate a calibration model even in highly complex mixture by excluding irrelevant information and spectral noise.

The Raman spectral features of paraffin, naphthene, and aromatic hydrocarbons were clearly different over the entire Raman region, however spectrum of naphtha was completely overlapped with each individual component in naphtha. PLS regression was utilized to resolve this complex mixture and quantify total PNA composition. The results showed good correlation with reference GC method and feasibility of Raman spectroscopy in conjunction with multivariate calibration technique as a quantitative analytical method, even in complex samples.

Experimental Section

Raman Spectra. FT-Raman spectra were collected on a Bruker IFS55 FTIR spectrometer equipped with a Bruker FRA 106 Raman module, CaF_2 beam splitter, and cryogenically cooled Ge detector. Both Stokes and anti-Stokes spectral region were collected. The excitation source used was a diode pumped Nd : YAG laser operated at 1064 nm and power of 200 mW. The Rayleigh line was rejected by using a dielectric filter disk. The sample was transferred in a rectangular cuvette with back-reflecting mirror and 128 scans were taken at a resolution of 4 cm^{-1} in a 180° scattering arrangement. Prior to Fourier transformation, the interferograms were apodized with a Blackman-Harris four-point filter and zero-filled. All the Raman spectra were collected at room temperature (23 ^{o}C).

Sample Preparation. Fifty naphtha samples were obtained over a 6 month period at SK Corporation (Ulsan, Korea). Over long period, samples were cautiously collected to give more concentration and process related variations into the data set that would result in more robust calibration models. The concentration variations of each component in

* To whom correspondence should be sent. Email: hoeil@ skcorp.com   Phone: +82-52-270-1301. Fax: +82-52-270-1309

Table 1. The concentration variations of each component in the data set

	Total Paraffin (vol%)	Total Naphthene (vol%)	Total Aromatic (vol%)
Maximum	60.81	38.24	17.82
Minimum	43.84	26.34	11.43
Average	54.67	31.30	13.95

the data set are summarized in Table 1. As shown in Table 1, the average concentrations of each component are varying from 13.95 to 54.67 vol %. Immediately after collection, samples were sealed and stored in a refrigerator at 4 °C to prevent evaporation of the hydrocarbons. The concentration of each sample was measured by a Hewlett-Packard 5890 GC (Reformulyzer™) equipped with three separation columns and a flame ionization detector (FID). Initially, the first column (OV-275, 3 m×3.1 mm i.d.) was utilized to separate naphtha to aromatic and non-aromatic hydrocarbons. Then pre-separated aromatic and non-aromatic hydrocarbons were transferred to OV-101 column (dimethylpolysiloxane, 4 m×3.1 mm i.d.) and 13X molecular sieve column (1.55 m×2 mm i.d.), respectively, where separation by carbon number was achieved. For analysis, 0.2 μL of naphtha sample was injected using a syringe and helium was used as a carrier gas. The temperature conditions of oven, injection port, and detector were 130, 120, and 150 °C, respectively.

PLS Regression. PLS regression has been performed using GRAMS/32 software with add-on PLS algorithm (Galactic Industries Corporation, Salem, NH). Raman spectra were imported into GRAMS/32 before the PLS regression was accomplished. A total of 50 spectra from 50 naphtha samples were divided into 35 spectra for calibration and 15 spectra for the prediction data set serving as a validation set. Spectra in the calibration and prediction set were randomly chosen.

Results and Discussion

Raman Spectral Features. For the selective and simultaneous quantification of components in a complex mixture like naphtha using PLS regression, it is required that the minute spectral differences among components should be recognized. For qualitative examination of spectral features of typical representative components in naphtha, Raman spectra of *n*-hexane, cyclohexane, and toluene are shown in Figure 1. All spectra were linearly offset for the purpose of qualitative comparison. In Raman spectra, very clear and sharp spectral differences among the components are observed and most useful information is located at 3200-2500 and 1700-300 cm^{-1} regions. For toluene, intense and sharp C-H bands attached to the benzene ring, which are noticeably different from aliphatic C-H bands, are clearly observed at 3050 and 1650 cm^{-1}. In addition, spectral features in 1700 to 300 cm^{-1} region are clearly different from those of aliphatic hydrocarbons. For cyclohexane, overall
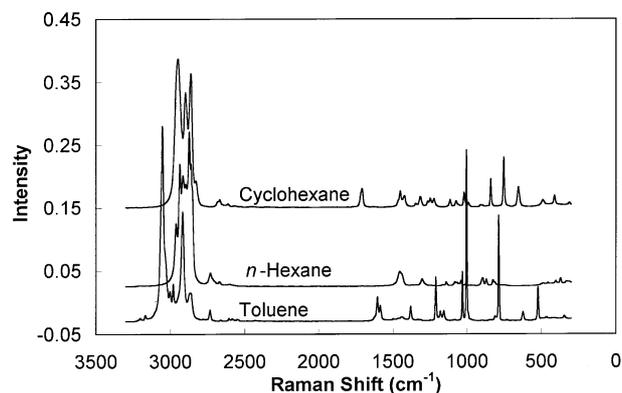
Figure 1. Raman spectra of *n*-hexane, cyclohexane, and toluene (offset for clarity).

spectral features are unique and different, especially the band at 750 cm^{-1} from the breathing motion of the hydrocarbon ring. The spectral features of *n*-hexane are quite different from other components. The overall Raman spectral features of each component are very clear and easily identified.

When Raman spectroscopy is used for the quantitative analysis of very complex mixtures such as naphtha, which is composed of more than one hundred individual components and various kinds of hydrocarbons with similar structures, the resulting spectrum will be highly overlapped regardless of spectral bandwidth. Raman spectrum of typical naphtha is shown in Figure 2 and the resulting spectrum is overlapped with each of the components in naphtha. Especially from the 1700 to 690 cm^{-1} range in the Raman spectrum, the sharp peaks of individual components that were shown in Figure 1 are overlapped and the resulting peaks are broad, relatively. The aromatic C-H band at 3050 cm^{-1} is still clearly identified. Raman spectrum of naphtha is highly overlapped, therefore multivariate calibration methods are essentially required to resolve this mixture.

PLS Calibration Models. In the application of the PLS algorithm, it is generally known that the spectral range and number of PLS factors are critical parameters. It has been previously determined that the calibration performance depends on the spectral range utilized.[10] Six different spec-
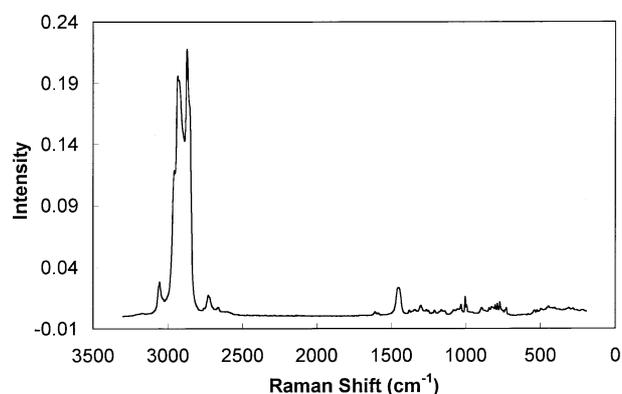
Figure 2. Raman spectrum of typical naphtha.
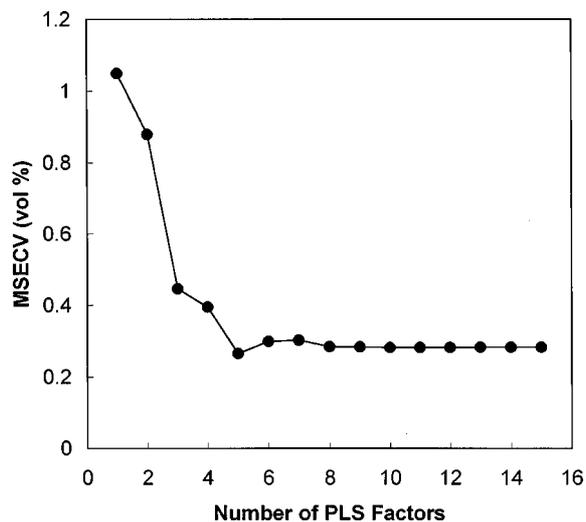
Figure 3. The effect of number of PLS factors on MSECV for total paraffin content determination within the 3200-2500 cm^{-1} range.

tral ranges of 3200-190, 3200-2500, 1700-190, 1700-670, 670-190 and the combination of 3200-2500 and 1700-190 cm^{-1} were examined. These spectral ranges were selected to include combinations of the three major scattering regions (3200-2500, 1700-670, and 670-190 cm^{-1}). The 3200-190 cm^{-1} range corresponds to the full spectral range without excluding non-scattering and noisy region of the 2500-1700 cm^{-1} range

   The optimum number of factors was identified as the number of factors that gives a minimum MSECV (Mean Square Error of Cross Validation). The cross validation was used as a validation method by dividing calibration data set into 7 segments. The MSECV is a good index to find the optimal number of PLS factors without overmodelling. Figure 3 shows MSECV plotted as functions of the number of PLS factors used for the determination of total paraffin content within the 3200-2500 cm^{-1}. As expected, the MSECV decreases sharply with the initial factors and gradually decrease as more paraffin-dependent spectral variation is incorporated into the calibration model. As the number of factor increase further, MSECV begins to increase at 6 PLS factors. An increase in the MSECV indicates that the data has been over-fitted by incorporating spectral information
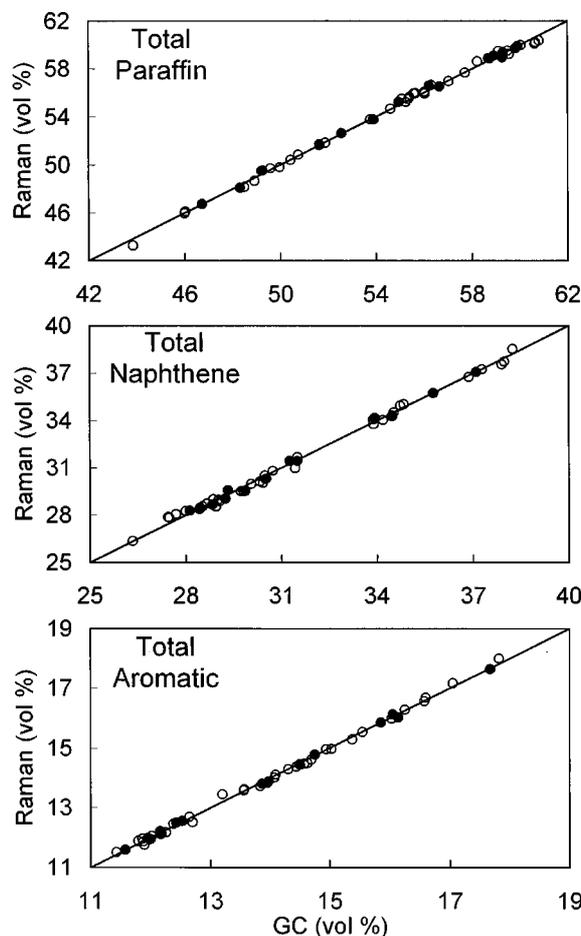
Figure 4. Scatter plots showing correlation between Raman and GC analyses for total paraffin (top), naphthene (middle), and aromatic (bottom) content using both 3200-2500 and 1700-190 cm^{-1} ranges. Open and filled circles correspond to calibration and prediction data, respectively.

into the model that is not related to paraffin. In this case, 5 PLS factors were chosen as the optimum. The optimal number of PLS factors for other spectral ranges and other components was determined by generating and examining the same type of plot.

   The calibration results for each component are summarized in Table 2. The optimal number of factors, MSECV,

Table 2. PLS calibration results for each component at different spectral range

		3200-2500 & 1700-190 cm^{-1}	3200-2500 cm^{-1}	1700-190 cm^{-1}	1700-670 cm^{-1}	670-190 cm^{-1}	3200-190 cm^{-1}
Total	Factors	5	5	7	6	3	5
Paraffin	MSECV (vol%)	0.27	0.27	0.74	0.87	0.58	0.28
	SEP (vol%)	0.17	0.18	0.38	0.44	0.53	0.17
Total	Factors	4	4	7	7	5	4
Naphthene	MSECV (vol%)	0.24	0.23	0.76	0.86	0.80	0.25
	SEP (vol%)	0.18	0.19	0.35	0.43	0.61	0.18
Total	Factors	6	6	4	4	7	6
Aromatic	MSECV (vol%)	0.11	0.10	0.16	0.15	0.40	0.12
	SEP (vol%)	0.06	0.06	0.10	0.11	0.26	0.07

and SEP (Standard Error of Prediction) of each component at different spectral ranges are shown in this table. SEP was calculated by predicting the samples in the prediction data set. In overall, the best results were achieved using 3200-190, 3200-2500, and the combination of 3200-2500 and 1700-190 cm^{-1}. The corresponding MSECVs and SEPs are similar each other. The inclusion of 3200-2500 cm^{-1} range in the calibration models resulted in the better calibration performance. As shown in Figure 2, Raman intensity is much higher in the 3200-2500 cm^{-1} range compared to that of the 1700-190 cm^{-1} range. Therefore the Raman intensity is much more sensitive to the concentration variations and signal-to-noise ratio is relatively higher in this range. Consequently the 3200-2500 cm^{-1} range, higher sensitivity and signal-to-noise ratio, provided the better calibration results. Using the 1700-190 cm^{-1} range solely or combinations including this range resulted in worse calibration results.

The calibration results of total aromatic content are better than those of total paraffin and naphthene content. As shown in Figure 1 and 2, unique and non-overlapping band of aromatic compounds at 3050 cm^{-1}, which is noticeably different to aliphatic C-H bands, gives good calibration performance. The calibration results of total paraffin and naphthene are similar.

Scatter plots showing the correlation between NIR and GC analyses for the calibration models of each component using both 3200-2500 and 1700-190 cm^{-1} ranges are shown in Figure 4. Open circles and filled circles represent the calibration and prediction data, respectively. In all cases, the calibration and prediction data has good correlation with conventional GC data and many points fall on or close to the unity line. Concentrations in the prediction data set were accurately predicted with SEP of 0.17 vol% for total paraffin, 0.18 vol% for total naphthene, and 0.06 vol% for total aromatic content.

Conclusion

The feasibility of Raman spectroscopy for the quantitative analysis of complex sample has been sufficiently demonstrated. The total paraffin, naphthene, and aromatic contents in naphtha can be successfully measured by Raman spectroscopy. The prediction results showed good correlation with those of GC. Using Raman spectroscopy, spectral differences among paraffin, naphthene, and aromatic hydrocarbons can provide the enough quantitative information for the calibration using PLS regression. Along with near-infrared (NIR) and infrared (IR) spectroscopy, FT-Raman spectroscopy has a strong potential for on-line quantitative analysis since it provides richer chemical information that can be easily interfaced with optical fiber for remote monitoring without sample pretreatment. Considerable efforts are still required to improve the signal-to-noise ratio so that the information of Raman spectroscopy have the precision and reproducibility needed. Future work will evaluate remote FT-Raman spectroscopy using optical fibers with a proper external reference for the measurement important properties in refinery and petrochemical areas.

References

1. Chung, W. M.; Wang, Q.; Sezerman, U.; Clarke, R. H. *Appl. Spectrosc.* **1991**, *45*, 1527.
2. Clarke, R. H.; Chung, W. M.; Wang, Q. S.; DeJesus, Sezerman, U. *J. Raman Spectrosc.* **1991**, *22*, 79.
3. Cooper, J. B.; Flecher, P. E.; Vess, T. M.; Welch, W. T. *Appl. Spectrosc.* **1995**, *49*, 586.
4. Martens, H.; Naes, T. M. *Multivariate Calibration*; John Wiley and Sons: New York, U. S. A., 1989.
5. Hendra, P.; Jones, C.; Warnes, G. *Fourier Transform Raman Spectroscopy, Instrumentation and Chemical Applications*; Ellis Horwood Ltd.: U. K., 1991.
6. Seasholtz, M. B.; Archibald, D. D.; Lorbe, A.; Kowalski, B. R. *Appl. Spectrosc.* **1989**, *43*, 1067.
7. Cooper, J. B.; Wise, K. L.; Jensen, B. J. *Anal. Chem.* **1997**, *69*, 1973.
8. Cooper, J. B.; Wise, K. L.; Groves, J.; Welch, W. T. *Anal. Chem.* **1995**, *67*, 4096.
9. de Bakker C. J.; Fredericks, P. M. *Appl. Spectrosc.* **1995**, *49*, 1766.
10. Chung, H.; Lee, J. S.; Ku, M. S. *Appl. Spectrosc.* **1998**, *52*, 885.