Anal. Chem. **2001,** *73,* 5677−5682

Two-Dimensional Gas Chromatography and Trilinear Partial Least Squares for the Quantitative Analysis of Aromatic and Naphthene Content in Naphtha

Bryan J. Prazen,[†] Kevin J. Johnson,[†] Andrew Weber,[‡] and Robert E. Synovec*,[†]

Center for Process Analytical Chemistry, Department of Chemistry, BG-10, Box 351700, University of Washington, Seattle, Washington 98195-1700, and The Dow Chemical Company, 2831 North Brazosport Boulevard, Freeport, Texas 77541.

Quantitative analysis of naphtha samples is demonstrated using comprehensive two-dimensional gas chromatography (GC × GC) and chemometrics. This work is aimed at providing a GC system for the quantitative and qualitative analysis of complex process streams for process monitoring and control. The high-speed GC × GC analysis of naphtha is accomplished through short GC columns, high carrier gas velocities, and partial chromatographic peak resolution followed by multivariate quantitative analysis. Six min GC × GC separations are analyzed with trilinear partial least squares (tri-PLS) to predict the aromatic and naphthene (cycloalkanes) content of naphtha samples. The 6-min GC × GC separation time is over 16 times faster than a single-GC-column standard method in which a single-column separation resolves the aromatic and naphthene compounds in naphtha and predicts the aromatic and naphthene percent concentrations through addition of the resolved signals. Acceptable quantitative precision is provided by GC × GC/tri-PLS.

Naphtha serves as an example of a complex chemical mixture to demonstrate the utility of high-speed, comprehensive two-dimensional gas chromatography (GC × GC) coupled to trilinear partial least squares (tri-PLS) data analysis. Naphtha is a petroleum distillation fraction typically ranging from propane to tetradecane. The distribution can vary somewhat, depending on the distillation cut. Naphtha can contain straight-chain (paraffins), branched-chain (isoparaffins), and cyclic (naphthenes) saturated and unsaturated (olefins) alkanes. Aromatic compounds such as benzene, toluene, xylenes, and ethylbenezne can be present in the percent levels. Typically, compounds are divided into the classes of paraffins, isoparaffins, aromatics, naphthenes, and olefins.

Gas chromatography (GC) is an effective analytical approach for volatile and semivolatile organic samples such as naphtha; however, complex samples, such as fuels, often require long GC analysis times. To reduce analysis times and broaden applicability, many GC separations are done using multicolumn systems in which columns of complementary selectivity are coupled. Conventional multicolumn chromatography using heart-cutting is a powerful technique, but only a fraction of the chemical components eluting from the first column are injected onto the second column for further separation.[1] The majority of the sample components are discarded or detected after a single-column separation. Although heart-cutting techniques provide high-resolution separation of some target compounds in the sample, they are not comprehensive, two-dimensional separations. The requirements for performing comprehensive two-dimensional (2-D) chromatographic separations have been defined in practice and recently reviewed.[2−4] Comprehensive two-dimensional chromatographic techniques, such as GC × GC, are well-suited for the analysis of complex mixtures, such as middle-range petroleum distillates such as naphtha.[5−9] Heart-cutting and one-dimensional (1-D)-column techniques lack the separation power for isolating complex mixture components quickly. They produce a 1-D retention vector that has insufficient analyte peak capacity per time to separate all or many components in an acceptable time. In contrast, comprehensive 2-D techniques are specifically designed to produce a planar separation space that has more room in which components may be distributed. A comprehensive 2-D separation is capable of separating more components of a complex mixture in less time than a 1-D separation.[10] Thus, GC × GC is rapidly emerging as a tool to analyze complex samples.

Powerful chemometrics techniques can be used to exploit the extra information contained in the 2-D data generated from GC

[†] University of Washington.
[‡] The Dow Chemical Co.

(1) Bertsch, W. In *Multidimensional Chromatography, Techniques and Applications*; Cortes, H. J., Ed.; Marcel Dekker: New York, 1990; Vol. 50, pp 74−144.
(2) Liu, Z.; Lee, M. L. *J. Microcolumn Sep.* **2000**, *12*, 241−254.
(3) Bushey, M. M.; Jorgenson, J. W. *Anal. Chem.* **1990**, *62*, 161−167.
(4) Bushey, M. M.; Jorgenson, J. W. *Anal. Chem.* **1990**, *62*, 978−984.
(5) Beens, J.; Blomberg, J.; Schoenmakers, P. J. *J. High Resolut. Chromatogr.* **2000**, *23*, 182−188.
(6) Beens, J.; Boelens, H.; Tijssen, R. *J. High Resolut. Chromatogr.* **1998**, *21*, 47−54.
(7) Frysinger, G. S.; Gaines, R. B. *J. Sep. Sci.* **2001**, *24*, 87−96.
(8) Schoenmakers, P. J.; Oomen, J. L. M. M.; Blomberg, J.; Genuit, W.; Velzen, G. V. *J. Chromatogr. A* **2000**, *892*, 29−46.
(9) van Deursen, M.; Beens, J.; Reijenga, J.; Lipman, P.; Cramers, C. *J. High Resolut. Chromatogr.* **2000**, *23*, 507−510.
(10) Fraga, C. G.; Bruckner, C. A.; Synovec, R. E. *Anal. Chem.* **2001**, *73*, 675−683.

10.1021/ac010637g CCC: $20.00
Published on Web 10/27/2001

× GC, provided the data fit a bilinear structure. Valve-based instruments, such as the one we have been developing, have been shown to provide bilinear data.[11−13] In previous work, target analytes in bilinear GC × GC data were successfully quantified using the chemometric technique known as generalized rank annihilation method (GRAM).[11−13] GRAM has the unique and powerful ability to quantify unresolved signals using only a single standard sample.[11,12,14−16] Unfortunately, using a single standard sample can also be a restriction. When multiple standards are used, a signal-to-noise advantage improves the calibration model. A potentially powerful approach to analyzing GC × GC data is tri-PLS.[17−21] Tri-PLS is an extension of partial least squares (PLS)[22] to instruments that form a two-dimensional matrix of data for each sample. Tri-, or three-way, refers to the data structure made when multiple two-dimensional data sets are stacked to obtain a three-dimensional matrix for the purpose of comparing data sets.

Producing a 2-D separation space with separation columns requires coupling two columns of different selectivity such that all components of a sample are subject to two different separations without either separation nullifying the other.[23−25] Figure 1 is a schematic representation of a GC × GC analyzer. The GC × GC analyzer is designed to produce a 2-D separation space using two GC columns. As sample components leave the first column, they are injected onto the second column by a computer-programmed, high-speed diaphragm valve. The rate of second-column injections is variable. In this work, the injections onto the second column occurred at a frequency of 1.33 Hz (750 ms second-column run time) in order to satisfactorily sample the narrowest peaks departing the first column (e.g., 3 s wide at the base), thus satisfying the criteria for a comprehensive 2-D separation.[2] The first column is usually a relatively larger bore, nonpolar capillary column that separates components primarily based on volatility. The second column is a high-speed, narrower-bore polar capillary column that, in this case, separates components based on dipole and hydrogen-bonding interactions.[26] The outlet of the second column is connected to a flame ionization detector (FID). The diaphragm valve used in this work repeatedly injects small portions of the first-column eluent into the second column. Most GC × GC experiments use thermal

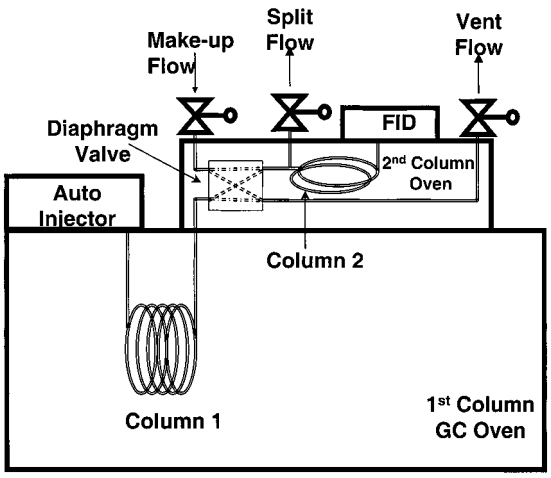

Figure 1. Schematic representation of dual oven GC × GC system. Compounds eluting from the nonpolar first column are diverted by a diaphragm valve to a polar second column prior to flame ionization detection. Dual ovens allow for independent temperature control of the two columns.

focusing techniques, either heating or cooling, to produce sample injections onto the second column.[27] In either case, each injection onto the second column generates a high-speed, secondary chromatogram. Each secondary separation needs to be fast enough so that multiple secondary chromatograms are generated across the width of a peak eluting from the first column.[2,28] A series of secondary chromatograms are arranged side-by-side to make a matrix of data that is a 2-D chromatogram with peaks dispersed over a 2-D space.

The valve-based GC × GC instrument has heretofore not been constructed in such a manner as to allow independent temperature control of both capillary columns.[11−13,26] Isothermal operation limits the range of samples that can be analyzed in a timely fashion. Furthermore, simultaneous temperature programming of both columns results in a less than orthogonal 2-D separation.[29] The GC × GC used in these experiments has the two columns in individual ovens, thus allowing for independent temperature control.[30] Independent temperature control allows for the high-speed analysis of a broad range of sample types. In these experiments, the temperature of the first column, which separates primarily based on volatility, was ramped, and the temperature of the second column, which separates primarily by dipole interactions, polarizability, and hydrogen-bond acidity interactions, is held constant. This allows compounds of a wide range of volatilities to be eluted from the first column in a timely manner and with consistent peak width, while the second-column separations do not change within the elution of components off the first column. If both columns were ramped together, the first sample of a peak eluting off the first column that is injected onto the second column will be subject to a different second-column

(11) Bruckner, C. A.; Prazen, B. J.; Synovec, R. E. *Anal. Chem.* **1998**, *70*, 2796−2804.
(12) Fraga, C. G.; Prazen, B. J.; Synovec, R. E. *Anal. Chem.* **2000**, *72*, 4154−4162.
(13) Fraga, C. G.; Prazen, B. J.; Synovec, R. E. *J. High Resolut. Chromatogr.* **2000**, *3*, 215−224.
(14) Prazen, B. J.; Synovec, R. E.; Kowalski, B. R. *Anal. Chem.* **1998**, *70*, 218−225.
(15) Prazen, B. J.; Bruckner, C. A.; Synovec, R. E.; Kowalski, B. R. *J. Microcolumn Sep.* **1999**, *11*, 97−107.
(16) Prazen, B. J.; Bruckner, C. A.; Synovec, R. E.; Kowalski, B. R. *Anal. Chem.* **1999**, *71*, 1093−1099.
(17) Hasegawa, K.; Arakawa, M.; Funatsu, K. *Chemom. Intell. Lab. Syst.* **1999**, *47*, 33−40.
(18) Smilde, A. K. *J. Chemom.* **1997**, *11*, 367−377.
(19) Bro, R.; Heimdal, H. *Chemom. Intell. Lab. Syst.* **1996**, *34*, 85−102.
(20) Bro, R. *J. Chemom.* **1996**, *10*, 47−61.
(21) de Jong, S. *J. Chemom.* **1998**, *12*, 77−81.
(22) Kramer, R. *Chemometric Techniques for Quantitative Analysis*; Marcel Dekker: New York, 1998.
(23) Giddings, J. C. *Anal. Chem.* **1984**, *56*, 1258A−1270A.
(24) Giddings, J. C. *J. High Resolut. Chromatogr.* **1987**, *10*, 319−323.
(25) Giddings, J. C. *Unified Separation Science*; Wiley: New York, 1991.
(26) Seeley, J. V.; Kramp, F.; Hicks, C. J. *Anal. Chem.* **2000**, *72*, 4346−4352.
(27) Lee, A. L.; Lewis, A. C.; Bartle, K. D.; Quaid, J. B. M.; Marriott, P. J. *J. Microcolumn Sep.* **2000**, *12*, 187−193.
(28) Murphy, R. E.; Schure, M. R.; Foley, J. P. *Anal. Chem.* **1998**, *70*, 1585−1594.
(29) Venkatramani, C. J.; Xu, J.; Phillips, J. B. *Anal. Chem.* **1996**, *68*, 1486−92.
(30) Johnson, K. J.; Prazen, B. J.; Synovec, R. E. Presented at the 24th International Symposium on Capillary Chromatography and Electrophoresis; Las Vegas, NV, 2001: Poster 105.

separation-column than a sample of the same peak eluting off the first column later. This results in the breakdown of the bilinear structure of the data. GC × GC instruments that employ thermal modulation of the eluting components between the first and second columns may not produce bilinear data, as well, and this will be a concern because a bilinear structure aids in the multivariate resolution and quantification of GC × GC data.[10−13] Thus, the valve-based GC × GC described herein with independent temperature control provides bilinear data amenable to multivariate data analysis.

In this work, the quantitative predictions of the total percent weights of two classes of compounds in a complex natural product, namely aromatic and naphthene (cycloalkane) content in naphtha, are predicted by GC × GC/tri-PLS. PLS calibration models are well-suited to problems of this type, because unlike GRAM, PLS methods are able to model properties (e.g., percent weight of aromatics) arising from multiple GC × GC peaks. The GC × GC/tri-PLS method will be shown to provide adequate precision and accuracy, as compared to a single-column standard GC method, but the GC × GC/tri-PLS method is over 16 times faster than the standard method.

EXPERIMENTAL SECTION

Eleven neat samples of naphtha were analyzed with either three or four repeated separations of each sample. The GC × GC system was set up on a Hewlett-Packard 6890 gas chromatograph with flame ionization detection (FID) (Agilent Technologies, Wilmington, DE). Figure 1 is a diagram of this system. An autosampler (Agilent Technologies) injected 1-μL samples into a split/splitless injector at 175 °C, which operated in split mode with a split ratio of 2.0. The column oven temperature was held at 28 °C for 1.9 min, then increased at 35 °C/min to 90 °C where it is held constant for 2 min. The oven containing the second column was 70 °C. The detector was set at 175 °C. The first-column head pressure was 26.65 psi. Helium was the carrier gas. A PC running HP Chemstation rev. A.06.03 (Agilent Technologies) controlled the GC. The first column of the GC × GC system was a 5.8-m × 320-μm i.d. capillary column with a 0.25-μm poly(5% diphenyl/95% dimethyl siloxane) film (DB-5, J&W, Folsom, California). The second column was a 1-m × 200-μm i.d. column with a 0.2-μm poly(ethylene glycol) stationary phase (HP-Wax, Agilent Technologies). The diaphragm valve (DV-12, 6 port, Valco Instrument Co. Inc., Houston, TX) was actuated for 14 ms every 750 ms during a 6-min GC × GC run. Thus, the valve repeatedly diverted a small portion of the first column effluent toward the second column. The effluent from the first column was split after the diaphragm valve between the second column and 0.4 m of 180-μm i.d. silica tubing. A large portion of the diaphragm-valve GC × GC system that was used in these experiments was previously described in detail.[11−13]

The FID signal from the GC × GC was amplified by an electrometer built in-house. The signal from the electrometer was measured at a rate of 100 000 points/s by a data acquisition board (model AT-MIO-16XE−50, National Instruments, Austin, TX) connected to a PC running LabVIEW 5.0 (National Instruments). The raw data was boxcar-averaged to 500 points/s and transferred into MATLAB 5.2 (The Mathworks Inc, Natick, MA), where it was converted into a matrix of data. Each row of the matrix from a GC × GC run represented a fixed time on the second GC column separation-time axis, and each column of the matrix represented a fixed time on the first GC column separation-time axis. The tri-PLS algorithm was implemented in MATLAB, and is part of the N-way toolbox (version 1.04). The N-way toolbox is available on the worldwide web.[31]

A standard single-column GC method providing a separation time in excess of 100 min was used to validate and calibrate the GC × GC quantification. In this method, 1-μL samples were injected using a split injector onto a Petrocol DH column that was 100 m × 0.25 mm with a 0.5-μm stationary phase (J &W Scientific, Folsom, CA). Hydrogen with a head pressure of 51 psi was the carrier gas. Flame ionization detection was used. The column oven temperature was held at 35 °C for 0.7 min, then it was increased at 1 °C/min to 60 °C, where it was held constant for 15 min, after which it was increased at 19 °C/min to 275 °C, where it was held constant for 65 min. Analytes were fully resolved, and quantification was performed using peak area. A complex automated procedure identified the peaks using a priori information and sorted them according to compound classes, such as aromatic and naphthenes. Signals of individual peaks were added to determine percent weight of aromatics and naphthenes.

RESULTS AND DISCUSSION

Raw data from the GC × GC analysis of naphtha is shown in Figure 2A. The time-dependent response of the FID from the GC × GC is in the form of a single vector that contains the modulation of consecutive second-column GC separations. Figure 2B is a separation of naphtha using a 100-min single-column GC method. Single-column separations such as the one in Figure 2B were used for validation and calibration of the GC × GC/tri-PLS quantification. This single-column analysis method will be referred to as the standard method. A comparison of parts A and B of Figure 2 demonstrates that when viewing the entire chromatogram, the raw GC × GC data resembles that of the single-column separation. From a distance, the detail of the second-column separation is not seen, but the modulation of the first-column separation is seen in the raw data. Note that the separation time for the GC × GC data shown in Figure 2A is only 6 min, whereas the separation time for the single-column method is 100 min. The combination of multivariate data analysis and the added selectivity of the second-column separation of the GC × GC allows for a 16-fold reduction in analysis time from that of the single-GC-column standard method.

Figure 3 demonstrates the selectivity advantage of the second-column separation in the GC × GC. The two peaks shown in Figure 3 are overlapped on the first-column separation at a retention time of 1.35 min, yet they are fully resolved in the second-column separation. The second-column separation time must remain short to retain first-column chromatographic information. Thus, fast second-column injections are necessary to create a high-speed GC × GC separation. In these experiments, a high-speed diaphragm valve is used to make injections onto the second column. A diaphragm valve is a simple yet effective means for this injection,[11−13,26] yet thermal modulation techniques are the most common means of GC × GC second-column injection.[32−36] The more retained peak of the two shown in Figure 3 has a peak

(31) Bro, R.; Andersson, C. A. http://www.models.kvl.dk/source/nwaytoolbox/.
(32) Liu, Z.; Phillips, J. B. *J. Chromatogr. Sci.* **1991**, *29*, 227−231.

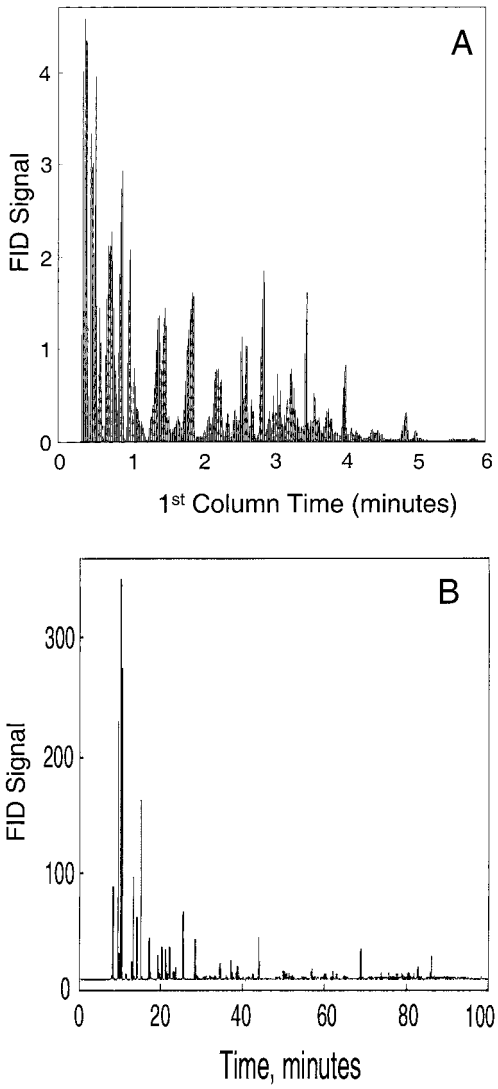

Figure 2. (A) Raw data from high-speed GC × GC naphtha separation prior to reformatting in the form of a matrix. (B) Single-column standard GC method naphtha separation.

width of ~40 ms. A recent comparison of second-column injection methods for GC × GC determined that the fastest of thermal modulation injectors result in peak widths of more than four times those shown in Figure 3.[27] Although diaphragm-valve-based systems do not send all of the effluent of the first column to the detector, this is only an important factor if the amount of sample available for injection onto the first column of the GC × GC is limiting the analysis.[12,13,26]

Raw GC × GC data sets such as the one shown in Figure 2A are transformed to two-dimensional matrix form by stacking consecutive second-column chromatograms. A contour map of the

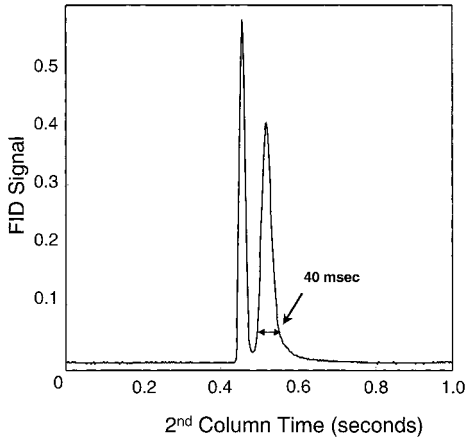

Figure 3. Second-column chromatogram at 1.35 min into the first-column separation of naphtha (see Figure 4). The two peaks are overlapped on the first-column separation but resolved on the second-column separation.

GC × GC separation of naphtha from Figure 2A is shown in Figure 4A. In this form, the matrix rows reflect the separation performed by the second GC column, and the matrix columns reflect the separation performed by the first column. The alkanes are not retained by the second column as much as the aromatics. Because the first- and second-column separations work differently upon unique compound classes, these classes are distributed on the 2-D separation "map" in identifiable patterns and locations. These compound class areas can simplify characterization and quantification, thus enhancing the power of the chromatographic information. Because of the limitations of the contour plot algorithm, many of the smaller peaks are not observed. Thus, in Figure 4B is shown a plot in which the locations of peak maxima are given. There are 75 peaks readily observed in Figure 4B for a typical sample. Clearly, one should not expect to observe as many peaks in a high-speed GC × GC run, as compared to the slower standard method as presented in Figure 2B. At first glance, it may seem that we are not significantly utilizing the 2-D separation power of GC × GC. However, even the relatively small amount of resolution on the second column of the GC × GC between many of the adjacent peaks is quite significant in the chemometric analysis by tri-PLS. The analytical approach presented here trades chromatographic resolution for saving run time and then using chemometrics to complete the analysis. The brackets in Figure 4A contain the portion of the GC × GC chromatogram used for tri-PLS quantification of aromatic compounds in the naphtha samples. Including readily available information about the approximate location of peaks of interest, such as the bracketed aromatic peaks, can improve the multivariate analysis results. By performing the analysis using only a portion of the GC × GC chromatograms, the effective signal-to-noise of the method is increased, because much of the data that do not include aromatic peaks is filtered out before the multivariate analysis. The tri-PLS analysis of naphthene compounds was performed using entire GC × GC chromatograms. Naphthenes are cycloalkanes and elute within the large band of peaks across the bottom of the GC × GC chromatogram.

Figure 5A,B shows the GC × GC/tri-PLS, quantitative, prediction results for naphtha samples. The quantitative results were

(33) Phillips, J. B.; Gaines, R. B.; Blomberg, J.; van der Wielen, F. W. M.; Dimandja, J. M.; Green, V.; Granger, J.; Patterson, D.; Racovalis, L.; de Geus, H. J.; de Boer, J.; Haglund, P.; Lipsky, J.; Sinha, V.; Ledford, E. B., Jr. *J. High Resolut. Chromatogr.* **1999**, *22*, 3−10.
(34) Kinghorn, R. M.; Marriott, P. J. *J. High Resolut. Chromatogr.* **1999**, *22*, 235−238.
(35) Blomberg, J.; Schoenmakers, P. J.; Beens, J.; Tijssen, R. *J. High Resolut. Chromatogr.* **1997**, *20*, 539−544.
(36) Frysinger, G. S.; Gaines, R. B.; Ledford, E. B., Jr. *J. High Resolut. Chromatogr.* **1999**, *22*, 195−200.

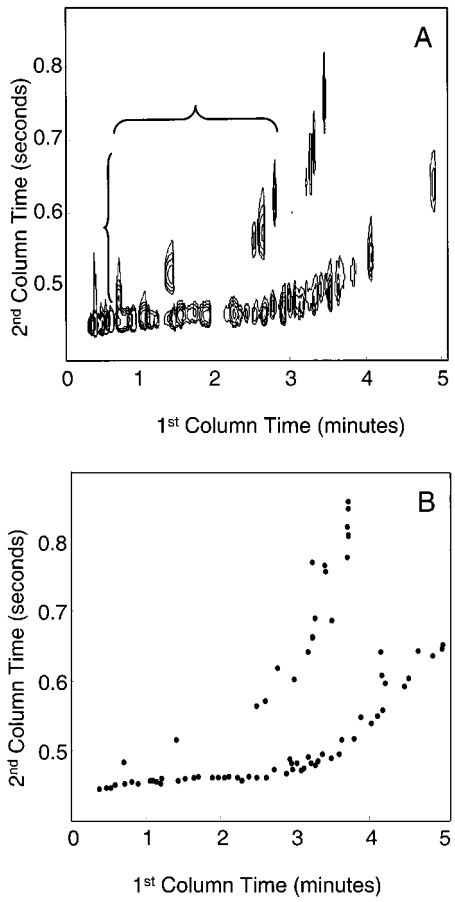

Figure 4. (A) Contour plot of data from the GC × GC analysis of naphtha. Accurate timing of the second-column injections allows for a simple transformation of the data in Figure 2A into this two-dimensional representation of the data. The brackets contain the portion of the data that was used for tri-PLS quantification of aromatic compounds in the naphtha samples. (B) Locations of peak maxima for the same GC × GC separation, allowing one to observe the smaller peaks not seen in A along with the larger peaks.

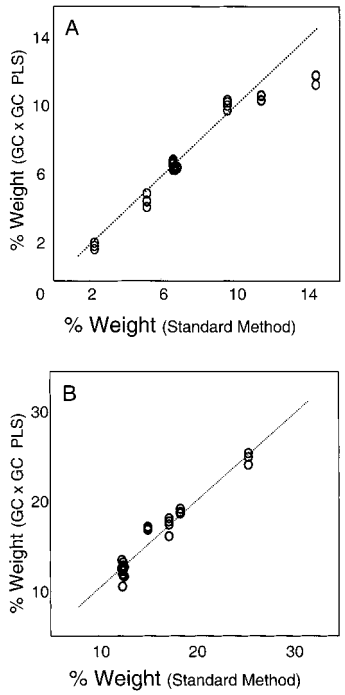

Figure 5. Quantitative results predicting percent weight by GC × GC/tri-PLS, as compared to that by the single-GC-column standard method for (A) aromatics and (B) naphthenes in naphtha.

determined by building a tri-PLS model using GC × GC data for all of the samples except the one being predicted, or leave-one-out cross-validation. Leave-one-out cross-validation was used for each analysis and gave an objective assessment of the prediction power of the methods while requiring only a small set of sample data. Replicate injections of each sample are shown in Figure 5A,B to demonstrate repeatability of this technique. The aromatic percent weight in the naphtha samples ranged from 2.25% to 14.65%. Unfortunately, the aromatic percent weights of many of the 11 samples available were ~6.6%. This makes it difficult for tri-PLS to form a robust model, yet most of the predictions were reasonably accurate. The sample with the highest aromatic concentration has a relatively poor prediction. This is because it was modeled using the remaining samples, and thus, calibration samples did not bracket the concentration of the sample being predicted. The rest of the results in Figure 5A indicate that a well-designed set of calibration samples that span the concentration range of future samples could lead to improvement in the accuracy of the tri-PLS aromatic percent weight predictions of high-percent-weight samples.[22] Figure 5B shows the percent weight of naphthenes in naphtha samples predicted using GC × GC/tri-PLS against the percent weight predicted using standard single-column GC. Although the high-speed GC × GC did not fully separate the naphthene compounds from the highly concentrated alkanes, good quantification was obtained for most samples using GC × GC/tri-PLS. The sample with the lowest percent weight (5.62%) and that with the highest percent weight (31.6%) did not predict well using GC × GC/tri-PLS and were, therefore, used only for tri-PLS calibration, not quantified themselves. Again, the less than desirable prediction of these two samples is attributed to the fact that percent weights of the calibration samples did not span that of the samples being predicted for these samples. All other samples were accurately modeled for naphthene content with the GC × GC/ tri-PLS system.

The role of the expert analyst must be minimized if GC × GC is to mature as an analytical method for routine on-line, at-line, or automated scenarios. Tri-PLS is an example of a robust multivariate analysis technique that could greatly reduce the time that is necessary to calibrate two-dimensional separations of complex samples. The standard method for aromatic and naphthene analysis of naphtha (see Figure 2B), which was used for calibration in the experiments for this paper, is a time-consuming single-column separation. Beyond the separation-time savings resulting from a change to GC × GC/tri-PLS analysis of naphtha, GC × GC/tri-PLS calibration has the potential to be much simpler than the standard method. In the standard method, an operator must accurately identify the retention time of 15 peaks each time separation conditions change. A GC × GC/tri-PLS calibration model can be built or updated by simply running the 6-min separation of approximately 10 known samples. Thus, automated calibration transfer is much easier with the GC × GC/tri-PLS analysis system. In practice, the two methods may be complementary. One may use an appropriate standard method to establish accuracy, and

then the high-speed GC × GC method may be employed in a rapid, automated fashion, as in on-line process analysis.

ACKNOWLEDGMENT

This work was supported by the Center for Process Analytical Chemistry (CPAC), a National Science Foundation-initiated University/Industry Cooperative Research Center at the University of Washington. The authors thank Carsten A. Bruckner for developing the software used to identify the peak maxima with various peak height ranges.

Received for review June 7, 2001. Accepted August 29, 2001.

AC010637G