

Coal Classification by HPLC and
Three-Dimensional Detection

Michael J. Sepaniak and Edward S. Yeung

Ames Laboratory* and Department of Chemistry
Iowa State University
Ames, Iowa 50011

Coal liquefaction products include a fraction known as the asphaltenes, which are operationally defined as the pentane insoluble, benzene soluble components. The asphaltenes have been postulated as intermediates in the conversion of coal to oil and contain a high concentration of the coal impurities. It is the goal of this work to characterize the asphaltenes derived from various solvent refined coals, so that similarities and differences between various refining processes can be identified. Such information can lead to the optimization of the processing conditions, improved quality control, reduced environmental pollution, and the understanding of the fundamental chemical reactions involved.

A typical coal-derived asphaltene sample contains hundreds of components, mostly in the 200-800 molecular weight range. Considering the complexity of these samples, it is imperative to resolve the components as much as possible, to obtain the maximum amount of information for classification purposes. Recent advances in microbore columns have resulted in extremely high efficiencies in liquid chromatography. Chromatographic runs however can last tens of hours. Also, the reliability, reproducibility, and useful life of these columns for repeated injections of such complex samples have not been adequately tested. We have therefore chosen for these studies standard commercial reversed-phase HPLC columns, which, with the proper eluant gradient, are not much lower in efficiency than those mentioned above. The use of a pre-column switching technique significantly prolonged column life and reduced contamination. The use of multidimensional detectors allows the extraction of a maximum amount of information, and the high sensitivity of the detectors allows small injection quantities and again more reliable separation.

Three optical detectors are used for this work, based on the fact that these are sensitive and that they complement one another in the type of information each provides. The first is a conventional uv absorption detector operating at 254 nm. This has a demonstrated detectability in the nanogram range and is a general and versatile detector. The second is a laser-excited fluorometric (LF) detector that has a detectability in the picogram range. In addition to the higher signal levels because of the higher photon fluxes in a laser, the monochromaticity in excitation results in correspondingly narrower Rayleigh and Raman lines and permits larger spectral windows for fluorescence observation. In order to obtain information as different from that in the first detector as possible, visible wavelengths are used in excitation and in emission. This favors the detection of the larger molecules, where conjugation shifts the absorption bands to longer wavelengths. The third is a two-photon excited fluorometric

*Operated for the U.S. Department of Energy by Iowa State University under contract No. W-7405-Eng-82. This research was supported by the Director for Energy Research, Office of Basic Energy Science, WPAS-KC-03-02-03.

(TPF) detector. A detectability in the nanogram range has been demonstrated with continuous-wave lasers, but picogram levels can be achieved if high-power pulsed lasers are used. The net excitation is into electronic states comparable in energy with those in uv absorption, but the unique selection rules of the TPF process provide complementary information. Furthermore, the TPF process is enhanced when a real electronic state matches the energy of one of the photons, in our case a visible photon. The selectivity again favors the larger molecules with a higher degree of conjugation.

Pattern recognition studies were performed using the subroutines CORREL, WEIGHT, BACLAS, SICLAS, MULTI, TREE, and KNN from the standard package known as ARTHUR. To limit the number of features in the statistical pattern recognition programs, an intra-sample feature correlation was first performed.

Reproducibility

Even though the injection quantities were kept to 40 μg , contamination still prevented the use of a single column for all the work, and 6 different columns (from the same production batch) were used. In the process of correlating the chromatographic peaks to extract features, we found that the normalized retention times and signal magnitudes among all injections had relative standard deviations of 2.5% and 9.2% respectively. Considering the complexity of these samples, the reproducibility is good.

Detector comparison

Chromatograms from each of the three detectors for a particular run are shown in Figure 1, indicating that information obtained from each is quite independent and the three complement one another. It should be noted that even if peaks occur at the same retention time, they need not correspond to the same component. A distinctive feature in Figure 1 is the positions of the different "center-of-mass", i.e., when about 50% of the weighted response has passed the detector in question. These are in the order uv, TPF and LF. The nature of reversed-phase separation using this gradient generally makes the smaller components elute early and the larger components elute late. Since the LF detector requires electronic conjugation for the necessary spectral red-shift, very little response shows up early in the chromatogram. Even when the individual concentrations of the components decrease towards the end of the chromatogram, as a result of our pre-column switching and as evidenced by the falling response in the uv detector, the LF response remains high. Chromatograms from the TPF detector present an interesting case. Presumably the abundance of two-photon states around 244 nm (twice the photon energy) is not too different from one-photon states at 254 nm (uv detector). The smaller response of TPF early in the chromatogram is an indication of the lower overall sensitivity of the process. The components towards the later part of the chromatograms are rich in electronic states in the 488 nm region, as seen from the LF chromatogram. These same electronic states serve to resonantly enhance the TPF process, thus providing a larger signal. The TPF still falls off a bit earlier than the LF signal, probably because of gradually decreasing concentrations and the lower overall sensitivity of the former process. Our choice of visible wavelengths in excitation and emission is thus justified, since the uv detector is not suitable for these components. Even if a uv fluorometric detector is used, one still favors the smaller, less conjugated components.

Sample comparison

Figure 2 shows chromatograms for the four different samples studied, all from the LF detector. Some consistent, distinctive features can be identified even without statistical analysis. The locations and the relative heights of the two major peaks in each chromatogram are sufficient to distinguish the PAMCO and the Cat. Inc. samples from the two Synthoil samples, and from each other. As expected, the two Synthoil samples are more difficult to distinguish. Minor features around 33 mins. and 37 mins. are useful for visual comparison in that case. At first sight it may seem that features after 50 mins. are distinctive for each sample. However, intra-sample correlation of these features is not good due to the pre-column switching procedure, and these features cannot be used.

A more objective comparison can be made using pattern recognition analysis. This was performed first for each detector with the 38, 38, and 41-dimensional vectors derived from the peaks via feature extraction. The results are presented in Table I. Then, the WEIGHT subroutine was used to determine the 6 or 7 most important features for each detector to the classification process. These were then used by themselves in a second pattern recognition analysis, the results of which are also shown in Table I. Finally, the combination of these 20 most important features and the 7 most important of these were also analyzed. It is significant to note that the combined set of 20 features did even better in every classification scheme than the individual sets of 38, 38, and 41 from the three detectors. In the combined set of 7 features, the success at classification is better than either the 6 TPF features or the 7 LF features, and about equal to the 7 uv features. This is consistent with the fact that in general having more independent types of measurements is better than having more of the same type of measurement for classification purposes. In Table II we list the overall ranking of the set of 20 combined features and the average retention times in the whole data set. The first observation is that the detectors are quite different, as expected. Features that are important in classification for one detector need not be important for another detector. This is seen from the relative locations of the selected features in the three chromatograms, and the lack of correlation among the individual rankings. One can infer that a distribution of molecular sizes and polarities can be important as features in classification, and that whole chromatograms rather than parts of chromatograms are needed. The second observation is that sometimes an important feature in one detector does correspond to an important feature in another detector. An example is the set of features UV2, LF3 and TPF2, which have indistinguishable retention times. Such situations magnify the significance of a given feature. Since the ultimate use of these classification and feature extraction studies is to identify compounds or classes of compounds that are important to the liquefaction processes themselves, these should be the most favorable starting points. Fractions can be collected on a semipreparatory scale in these regions, so that structural studies by infrared, nmr, or mass spectroscopy can be performed. Also, whenever the relative signals of a given feature at all three detectors remain constant in going from one sample to the next, more confidence can be put on the possibility that only a single component is involved. The third observation is that major peaks are not necessary useful features for classification. This points out the dangers of relying on just visual comparisons for classification. The human perception is easily biased towards the larger, seemingly better resolved, peaks.

In these complex samples, peaks may contain several components, so that slight differences in the gradient can dramatically alter the peak height. Also, if only one of the several components is important for classification, one would not be able to use it due to interference from the other components. The fourth observation is that the internal standard peak in the TPF chromatograms was actually identified as the eighth most important feature for classification. This apparent paradox can be resolved by noting that there is a continuous background at that location that is real. The "feature" therefore is simply the magnitude of this background at the retention time of the internal standard, not at all unreasonable to be important for classification. The fifth observation is that the six TPF features rank low in the group of 20 combined features for classification. This is more due to the poorer signal to noise levels and greater dependence on focusing and positioning, rather than an indication of the type of information. A higher response using e.g. pulsed lasers should improve things.

To assess the importance of the unresolved components that make up the broad background, we have extracted features based on the chromatographic valleys. The same pattern recognition calculations were performed using the signal levels in each detector at these "valleys". In general, the valleys are slightly poorer features for classification compared to the peaks. The fact that they perform this well indicates that the unresolved components can also be useful for characterization of the coal liquids. This is not surprising considering the diversity of components that are present, and that a large response in any detector need not be related to the importance of a given component. It is however incorrect to conclude that low-resolution chromatograms can be equally useful for classification purposes, since the success of these "valleys" is due in part to the reproducibility of the locations of the "peaks" that are around them. There are regions where both useful peaks and useful valleys are found, for example between 5 and 10 mins. in retention time. This then points to a particularly interesting region for further studies, possibly at higher chromatographic resolution and by collecting fractions for other analytical methods. The usefulness of the valleys in classification shows that the human bias towards recognizing peaks as features can be deceptive in complex systems.

Table I. Success at classification for each detector.

Detector	# Features	Success(%)			
		SICLAS	BACLAS	KNN	MULTI
TPF	38	90	100	42	100
TPF	6	74	94	52	87
LF	38	87	100	42	100
LF	7	90	94	74	94
UV	41	97	100	61	100
UV	7	90	100	90	100
All	20	100	100	94	100
All	7	97	100	84	100

Table II. Ranking of importance of peaks as features.

Rank	Feature*	Retention Time (min.)
1	LF1	21.1
2	UV1	26.6
3	LF2	8.1
4	LF3	40.5
5	LF4	26.0
6	LF5	6.3
7	UV2	40.3
8	LF6	9.1
9	UV3	21.8
10	TPF1	5.9
11	UV4	7.0
12	TPF2	40.0
13	UV5	19.7
14	UV6	5.3
15	TPF3	43.0
16	TPF4	52.5
17	TPF5	27.2
18	UV7	34.1
19	LF7	33.4
20	TPF6	24.4

*Features are identified in Figure 1. Letters refer to the respective detectors and numbers refer to those on the chromatograms.

FIGURE CAPTIONS

Figure 1. Information from the three detectors for the same injection. Numbers are peaks used as features as explained in the text. (a) uv detector; (b) LF detector; and (c) TPF detector.

Figure 2. Sample comparison using the LF detector. The labels refer to the four solvent-refined coals specified in the text.

FIGURE 1

FIGURE 2

