Improved resolution of hydrocarbon structures and constitutional isomers in complex mixtures using Gas Chromatography-Vacuum Ultraviolet-Mass Spectrometry (GC-VUV-MS)

Gabriel Isaacman[a], Kevin R. Wilson[b], Arthur W. H. Chan[a], David R. Worton[a,c], Joel R. Kimmel[d,e,f], Theodora Nah[g], Thorsten Hohaus[d], Marc Gonin[f], Jesse H. Kroll[h,i], Doug R. Worsnop[d], and Allen H. Goldstein[a,j,k]*

AUTHOR ADDRESS:

University of California, Berkeley

Department of Environmental Science, Policy, and Management

130 Mulford Hall #3114

Berkeley, CA 94720-3114

CORRESPONDING AUTHOR EMAIL ADDRESS: gabriel.isaacman@berkeley.edu

AUTHOR AFFILIATIONS:

[a] Environmental Science, Policy, and Management, University of California, Berkeley, CA

[b] Chemical Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA

[c] Aerosol Dynamics Inc., Berkeley, CA

[d] Center for Aerosol and Cloud Chemistry, Aerodyne Research, Inc., Billerica, MA

[e] Cooperative Institute for Research in the Environmental Sciences and Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO

[f] Tofwerk AG, Thun, Switzerland

[g] Chemistry, University of California, Berkeley, CA

[h] Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA

[i] Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA

[j] Environmental and Energy Technologies Division, Lawrence Berkeley National Laboratory, Berkeley, CA

[k] Civil and Environmental Engineering, University of California, Berkeley, CA

ABSTRACT Understanding the composition of complex hydrocarbon mixtures is important for environmental studies in a variety of fields, but many prevalent compounds cannot be confidently identified using traditional gas chromatography-mass spectrometry

(GC-MS) techniques. This work uses vacuum-ultraviolet (VUV) ionization to elucidate the structures of a traditionally "unresolved complex mixture" by separating components by GC retention time, t_R, and mass-to-charge ratio, m/Q, which are used to determine carbon number, N_C, and the number of rings and double bonds, N_{DBE}. Constitutional isomers are resolved based on t_R, enabling the most complete quantitative analysis to date of structural isomers in an environmentally-relevant hydrocarbon mixture. Unknown compounds are classified in this work by carbon number, degree of saturation, presence of rings, and degree of branching, providing structural constraints. The capabilities of this analysis are explored using diesel fuel, in which constitutional isomer distribution patterns are shown to be reproducible between carbon numbers and follow predictable rules. Nearly half of the aliphatic hydrocarbon mass is shown to be branched, suggesting branching is more important in diesel fuel than previously shown. The classification of unknown hydrocarbons and the resolution of constitutional isomers significantly improves resolution capabilities for any complex hydrocarbon mixture.

1. INTRODUCTION

Chemical analyses of environmental contamination often rely on chemical speciation of complex organic mixtures. Quantifying known tracer compounds in these environments can provide insight into biological and chemical degradation processes[1,2] and contaminant sources.[3-6] Over the past few decades, a variety of liquid and gas chromatography (LC and GC) techniques have been used with great success to improve understanding of pollutants as well as transformations in intricate natural systems. One of the newest promising techniques is vacuum ultraviolet (VUV) single-photon ionization mass spectrometry (SPI-MS).[7,8]

Traditional GC-MS employs electron impact (EI) ionization, where a molecule is ionized by an electron at an energy near the peak of ionization efficiency (usually 70 eV). This technique is broadly applicable, as nearly all organic compounds can be ionized at this energy, and generation of electrons is easily achieved using a filament in a vacuum chamber. However, because the ionization energy (IE) of most large organic compounds lies below 11 eV,[9] EI imparts considerable excess energy to the ionized molecule, causing dissociation into smaller fragment ions. For EI-MS analysis of pure compounds, parent ions are sometimes visible and/or the fragmentation pattern can be used to deduce molecular structure, but because EI of different organic compounds often yields similar fragment ions, interpretation of EI-MS of organic mixtures can be ambiguous. Alternatively, by using photons with energy of 11 eV or less, organic compounds can be ionized with minimal excess internal energy and therefore fragmentation, thus facilitating the identification of a compound by its molecular formula. This "soft" ionization is achieved in SPI-MS with ultraviolet photons in a vacuum (hence "vacuum-ultraviolet", or VUV) generated by a laser,[10] excitation of a rare gas,[11] or, in the case of this work, synchrotron radiation.[12] Soft ionization mass spectrometry has proven to be a valuable tool for analysis of complex mixtures because parent ions can be used for identification. This allows characterization in unprecedented detail when coupled to GC separation. For this reason, GC-VUV-MS analysis is being applied to a wide array of environmental and public health fields.[8]

Chromatography of many environmental samples yields a large "unresolved complex mixture" that has yet to be well quantified or described. This is because aliphatic hydrocarbons, a significant source of environmental contamination, fragment heavily

upon EI ionization, yielding many smaller fragments that are indistinguishable. Consequently large straight-chain alkanes do not yield significantly different mass spectra nor are they distinguishable from branched isomers. Though there is a small parent ion contribution to the mass spectra of many *n*-alkanes, it is not typically large enough to be used for unambiguous identification in a complex mixture. This parent ion is even smaller for branched alkanes, significantly hindering identification of branched hydrocarbons using EI ionization, though such compounds are common in environmental samples. VUV ionization allows improved identification of hydrocarbons because branched and straight-chain hydrocarbons both display a significant parent ion peak in photoionization mass spectra, although branched alkanes typically exhibit a greater degree of fragmentation. Examples of VUV mass spectra are shown in the Supplementary Information. This work uses the capabilities of GC-VUV-MS to develop methods for resolving the "unresolved complex mixture" by classifying compounds according to molecular mass and structure, including separation of structural isomers.

Applications of GC-VUV-MS analyses have often focused on petroleum fuels as sample organic mixtures[7,8,10,13,14] because of their environmental importance and daunting complexity. Diesel fuel serves as a useful example for typical pollution while its complexity provides an additional test of the resolution and capabilities of a separation technique. Diesel has also been analyzed in great detail using traditional GC-EI-MS techniques,[15] which is useful in comparing results from traditional methods to VUV ionization techniques.

GC-VUV-MS allows separation in the two-dimensional plane of chromatographic retention time (t_R, typically a function of volatility) and ion mass-to-charge ratio (m/Q).

Previous work[7,13] has effectively used this GCxMS plane to identify classes of compounds in complex mixtures by parent ion. Using this method, the volatility distribution of petroleum fuels was shown to approximately agree with GC-MS analysis of diesel fuel,[15] consisting primarily of saturated aliphatic hydrocarbons and polycyclic aromatic hydrocarbons (PAHs). However, the dependence of retention time on chemical structure allows GCxMS analysis to separate not only classes of compounds, but also many isomers. While this has been shown previously,[14] no attempt has been made to explore this capability in depth or determine the structures of unknown compounds. Owing to the limitations of GC-EI-MS in resolving complex hydrocarbon mixtures, only the most dominant branched hydrocarbons, such as pristane and phytane in diesel fuel, have been able to be quantified in past work.[15] Therefore, no complete picture of isomer content of petroleum mixtures is currently available. Isomer separation is a potentially important step forward in understanding fates of fossil fuel contaminants in the environment because structure of a compound strongly affects degradation and oxidation pathways,[16,17] which has implications for applications ranging from bioremediation to atmospheric aerosol modeling. Identification and quantification of constitutional isomers is useful beyond elucidating degradation pathways, providing insights into chemical ecology such as insect interactions.[18-21]

Using GC-VUV-MS, we improve the identification of typically unresolved compounds in complex hydrocarbon mixtures and demonstrate the quantitative resolution of structural isomers. Determination of carbon number, compound class, and branching provides structural information about any unknown compound in a mixture. These concepts and the optimization of VUV-MS for GC applications are shown in the context

of diesel fuel, as it is a useful surrogate for many environmental systems, but are intended to be applicable for a wide variety of systems. Distribution of isomers in a diesel sample is shown to be more predictable and constrained than might be expected by the number of potential isomers.

2. EXPERIMENTAL

2.1. Instrumentation

Samples and standards were analyzed by a GC (6890 Series, Agilent) coupled to a time-of-flight mass spectrometer (TOFMS; HTOF model, Tofwerk). Helium was used as a GC carrier gas, with a temperature ramp from 80 to 320 ºC. The effluent from the GC column (Rxi-5ms, 30 m x 0.25 mm x 0.25 µm, Restek Corporation) entered the TOFMS ionization chamber at vacuum. Molecules in this gas stream were then ionized either by electron impact (EI) ionization using an internal filament or vacuum-ultraviolet (VUV) photoionization at the Chemical Dynamics Beamline of the Advanced Light Source (ALS) at Lawrence Berkeley National Lab. The photon beam passes through a gas filter and MgF_2 window to remove unwanted undulator harmonics, resulting in photons that are tunable between approximately 8 and 11 eV. The heated region through which the GC column entered the TOFMS was varied between 150 and 275 ºC, as was the temperature of the ionization chamber (see Section 2.2. below for detailed discussion). The TOFMS was coupled to the beamline using custom modifications to the ionization chamber; the photon beam entered the ionization chamber perpendicular to the column outflow and the EI filament, allowing photoionization without removing the filament.

All standards were purchased from Sigma-Aldrich and AccuStandard. Diesel sample was purchased from AccuStandard. Data analysis was performed using custom processing code written in Igor Pro 6.0.4 (Wavemetrics).

2.2. Parameter optimization

Operating conditions can be adjusted by varying temperature and photon energy to minimize fragmentation; lowering these parameters increases the ion fraction observed as the parent ion (f_P) in a mass spectrum by decreasing internal energy.[22] Experimental conditions were optimized to find a balance between chromatographic separation and minimal fragmentation. In gas chromatography, the elution of compounds is dependent upon heating the chromatographic column using a temperature ramp that is optimized for speed and resolution so cannot be significantly modified. However, the temperature of the transfer line between the GC and the TOFMS, as well as the temperature of the ionization chamber in the TOFMS, can be varied; in this work, these temperatures were always modified together to create a uniform "transfer temperature." Using an injection of a standard containing C_8 through C_{40} *n*-alkanes, fragmentation was observed to be strongly dependent on transfer temperature (Figure 1). The gap between internal energy and IE widens as IE decreases with carbon number, causing the associated downward trend in f_P observed in Figure 1. Branched alkanes fragment approximately twice as much as straight-chain alkanes and are more sensitive to photon energy because they tend to have lower ionization energies.[22]

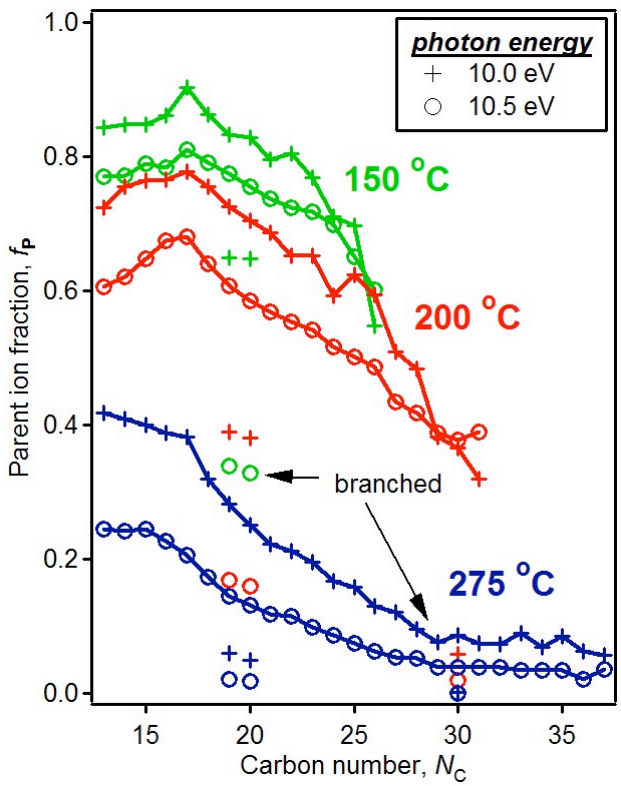


Figure 1. Fraction of ion abundance in the mass spectrum observed as the parent ion peak (f_P) for *n*-alkanes in the range of C_{14} to C_{37} at differing transfer temperatures (green: 150 ºC; red: 200 ºC; blue: 275 ºC) and photon energies (open circles: 10.5 eV; crosses 10.0 eV). Branched alkanes shown are pristane (C_{19}), phytane (C_{20}), and squalane (C_{30}).

In contrast to transfer temperature sensitivity, photon energy was observed to have a much smaller effect on f_P in the mass spectra of a variety of compound classes shown in the Supplementary Information, including acids, ketones, esters, and aromatics. Photon energy is an effective means of identifying unresolved compounds by selectively ionizing compounds that differ in ionization energies by more than 0.2 - 0.3 eV, as this is the width of the energy distribution of photons generated at the ALS.[23] Straight-chain alkanes in the C_{10}-C_{30} range are expected to have ionization energies at or near 10 eV, while the ionization energies of aromatics and other unsaturated hydrocarbons are expected to be significantly lower.[10,22,24] Photons of 10.5 eV were used for most purpose in this work to ionize all compounds of interest, but photons of 10.0 eV were used to selectively ionize unsaturated alkanes, resolving unsaturated and cyclic compounds of identical mass.

In Figure 1, lower volatility alkanes are not shown at low transfer temperatures because under these conditions they cannot be efficiently transferred to the TOFMS, demonstrating the balance between fragmentation and chromatographic transfer needs. The optimized parameters for any GC-VUV-MS application vary depending on analyte volatility range of interest. Diesel fuel, the hydrocarbon mixture used in this study, is not expected to contain significant organic mass in the range of C_{26}-C_{40}, so a transfer temperature of 150 ºC (green line in Figure 1) was used in these analyses..

2.3. Classification by double bond equivalents

In this work, the mass of a given hydrocarbon molecule, i, will be used to classify compounds according to carbon number, N_C, and number of double bonds and rings in

their structure, N_{DBE}. Each ring or double bond in the molecule i removes two H atoms compared to the saturated alkane $\mathrm{C}_{N_{\mathrm{C}}(i)}\mathrm{H}_{N_{\mathrm{C}}(i)+2}$, thereby reducing the mass of the hydrocarbon, $m(i)$, by twice the mass of hydrogen, $2*m(\mathrm{H})$. $N_{\mathrm{C}}(i)$ can reliably be estimated by the GCxMS characteristics of the molecule as discussed in Section 3.2. The mass of the saturated alkane therefore equals $m\left(\mathrm{C}_{N_{\mathrm{C}}(i)}\mathrm{H}_{N_{\mathrm{C}}(i)+2}\right)$. The number of double bonds and rings, called double bond equivalent number, N_{DBE}, becomes:

$$N_{\mathrm{DBE}}(i) = \frac{m\left(\mathrm{C}_{N_{\mathrm{C}}(i)}\mathrm{H}_{N_{\mathrm{C}}(i)+2}\right) - m(i)}{2 \cdot m(\mathrm{H})} \qquad \text{(Eq. 1)}$$

Geissler et al.[13] have shown that each N_{DBE} (adaptation of "class residue" from reference) is expected to represent a different class of compounds common in diesel fuel. The assignment of DBE classes in this work is shown in Table 1 and is based on fuel analysis performed by Schauer et al.,[15] Geissler et al.,[13] and Mitschke et al.[14]

N_{DBE}	Primary constituents
0	alkanes
1	cycloalkanes
2	bicycloalkanes
4	benzenes
7	naphthalenes
8	acenaphthenes
9+	higher PAHs

Table 1. Primary constituents of relevant double bond equivalent (DBE) classes

Unsaturated aliphatic hydrocarbons are not represented in Table 1 because they are not expected to be significant components of diesel fuel. For instance, although alkenes would be expected at $N_{DBE} = 1$, this DBE class is presumed in this and previous work to consist mostly of cycloalkanes. Accurate DBE classification relies on GC separation to distinguish between alkyl and PAH compounds of equal mass (e.g. $N_{DBE} = 0$ vs. $N_{DBE} = 7$), which is possible because chromatographic separation is a function of not only molecular formula, but also of chemical structure such as the presences of rings and branches.

3. RESULTS AND DISCUSSION

3.1. Spectra of known compounds

In the Supplementary Information, we provide the first compilation of VUV mass spectra of large organic compounds, with molecular masses primarily in the range of 150 to 400 Da. Spectra of 84 known compounds are shown, representing a wide range of functional groups: saturated and unsaturated aliphatic hydrocarbons including hopanes and steranes, polycyclic aromatic hydrocarbons, aliphatic and aromatic acids and esters, aliphatic ketones and aldehydes, oxygenated and multi-functional aromatics, and chloro- and nitro-aromatics. Spectra are collected using VUV ionization at 10.5 eV with transfer temperatures of 150 ºC (for a subset of 38 compounds) and 275 ºC (for all compounds). All spectra are collected by GC-VUV-MS of known compound mixtures.

From the spectra it is clear that, as stated above, transfer temperature has a significant impact on fragmentation. A transfer temperature of 150 ºC increases f_P for all compounds, while higher transfer temperatures enhance fragment peaks. Low

temperature operating conditions can therefore be used to focus on resolution by molecular mass as is done in this work. However, using high temperature operating conditions provides some structural information, potentially facilitating identification - as is done in EI - while still providing the molecular mass in most cases; these capabilities may be preferable for identification of compounds in some mixtures.

3.2. Classification by carbon number

An injection of diesel fuel (7.5 μg diluted to 1.5 μL with chloroform) was analyzed by GC-VUV-MS to explore resolution capabilities. The sample is shown in the GCxMS plane in Figure 2, where each chromatographic peak is shown as a circle, while the gray areas in the background represent ion abundance. Note that some DBE classes that do not appear to have a significant number of peaks contain long gray areas, indicating a rising background with no identifiable chromatographic peaks. Such areas are likely composed of many small chromatographic peaks that are not well-resolved but in some cases account for a non-negligible fraction of total detector response. It is clear from Figure 2 that certain DBE classes contain more peaks than others; it is these highly populated classes that are colored in Figure 2 and listed in Table 1. This work will focus primarily on the aliphatic DBE classes, as this best utilizes the soft ionization capabilities of VUV-MS; aromatics and PAHs tend not to fragment heavily even using EI ionization owing to resonance stabilization.

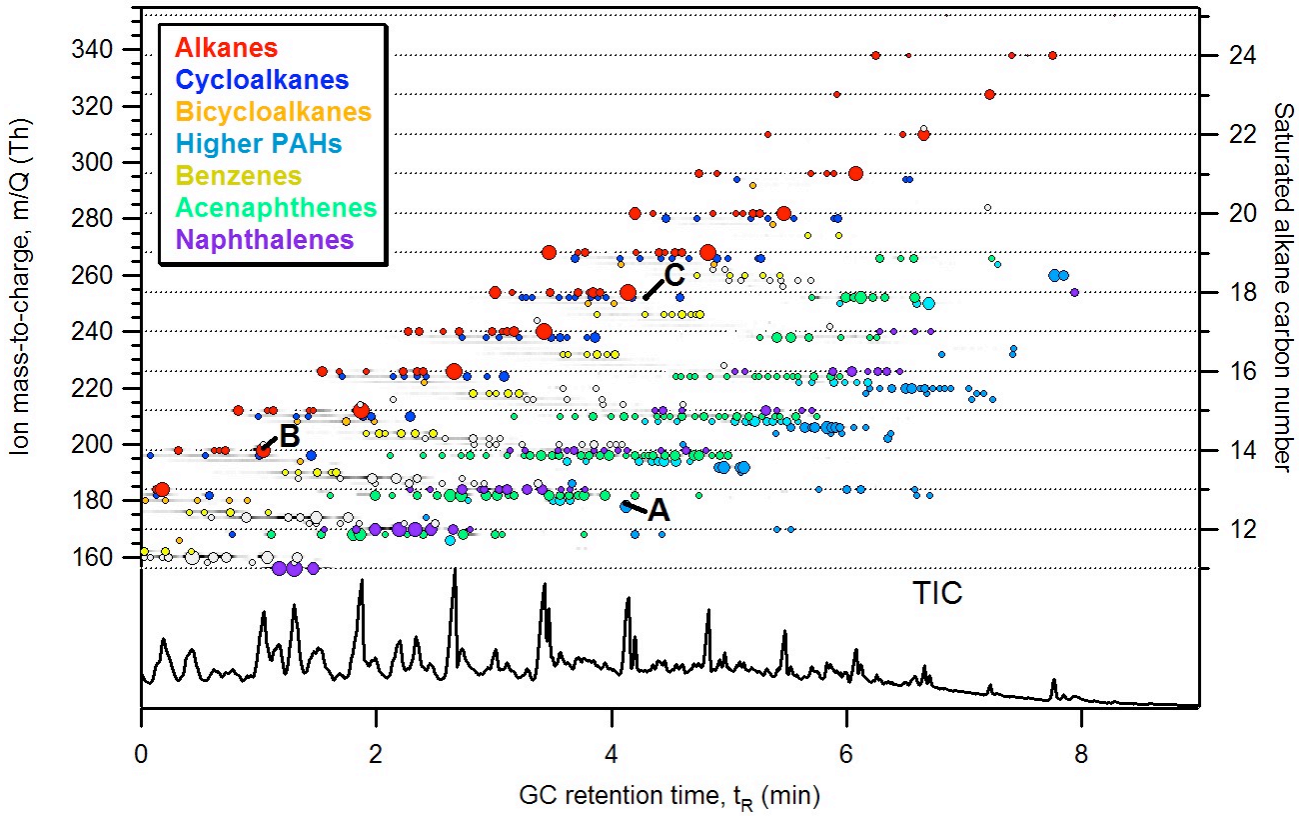


Figure 2. GCxMS plot of diesel fuel at a transfer temperature of 150 ºC and a photon energy of 10.5 eV. Circles represent peaks and are approximately sized by logarithmic peak height and colored by DBE class as assigned in Table 1 (white: unassigned class, red: alkanes, blue: cycloalkanes, orange: bicycloalkanes, yellow: benzenes, purple: napthalenes, green: acenaphthenes, blues: higher order PAHs). Gray areas in background are ion abundance, which is in some cases a large background hump with no resolved peaks. Dashed horizontal lines correspond to the mass of saturated alkanes and therefore $N_{DBE} = 0$; note that $N_{DBE} = 7$ for the following carbon number (naphthalenes) falls on this line as well. The total ion chromatogram (TIC) at the bottom shows total signal intensity; the largest peaks line up with the largest circles, which are *n*-alkanes. The compounds represented by the blue circle labeled "A" and the red circle labeled "B" are phenanthrene ($C_{14}H_{10}$) and tetradecane ($C_{14}H_{30}$) respectively and are discussed in Section 3.2. The

compounds represented by the blue circle labeled "C" is an unknown compound identified in Section 3.3 as a C_{18} cycloalkane with a single methyl branch.

The presences of rings and double bonds causes compounds in higher DBE classes to shift toward longer retention times, with aromatics eluting later than aliphatics with the same mass. Consequently, compounds of the same carbon number fall in diagonal regions in the GCxMS plane. A clear example of this phenomenon is illustrated by phenanthrene and tetradecane (compounds A and B, respectively, in Figure 2). Though phenanthrene, a C_{14} PAH with N_{DBE} = 10, has a retention time similar to octadecane ($C_{18}H_{38}$) and a mass equal to a C_{13} compound with N_{DBE} = 3, it falls into a diagonal band of peaks containing tetradecane ($C_{14}H_{30}$). These compounds therefore define an area of C_{14} compounds spanning DBE classes of 0 to approximately 11 (beyond which there are few peaks). There is an analogous band for all observed carbon numbers, so any peak or unknown compound can be classified by N_C and N_{DBE}.

Quantitation of the composition of diesel fuel by carbon number and DBE classifications (Figure 3) using VUV ionization is performed here as a validation against previous analyses of petroleum fuels. Signal intensity at each DBE and carbon number was corrected for transfer efficiency through the GC (which is approximately retention time dependent) and was then calibrated against signal intensity to one of three standards: hexadecane for aliphatic compounds (N_{DBE} = 0 through 3), dodecylbenzene for alkyl aromatics (N_{DBE} = 4 through 6), phenanthrene for PAHs (N_{DBE} >= 7). Absorption cross-section and detector response are expected to be structure and functional group dependent[10] so are neglected here by calibrating to compounds of similar structure. Compounds containing fewer than 14 carbons eluted during the chromatographic solvent delay so were not quantified in this work. Consequently smaller aromatics (such as C_2-alkylnaphthalenes and short-chain alkylbenzenes) as well as short-chain alkanes were not

measured. In this sample, alkanes account for nearly half (41%), while total aliphatic compounds constitute approximately three-quarters (73%) of the observed mass fraction. C_3-naphthalenes are expected to account for only approximately 10% of all naphthalenes[15] so it is likely that the importance of naphthalenes is under-represented in this analysis, though not by a factor of 10 as including lower carbon numbers would increase the observed mass of other DBE classes. Cycloalkanes comprise a more significant mass fraction (14%) of diesel fuel than has been shown in past work, though this may be a function of the specific sample of fuel analyzed. The compound distribution shown in Figure 3 approximately agrees with previous diesel fuel characterization[15] and suggests a volatility distribution that is similar to that measured by VUV-MS of crude oil.[13] This agreement with past techniques demonstrates that quantitation using GC-VUV-MS is a simple and effective method for composition analysis of hydrocarbon mixtures and classification by carbon number yields similar results to past work using other methods. Characterization by N_C and N_{DBE} may be useful in parameterization of chemically explicit models. Further study of the effects of branching on fragmentation, as well as calibrating with authentic standards for a wider variety of observed compounds, will improve future quantitation by GCxMS analysis.

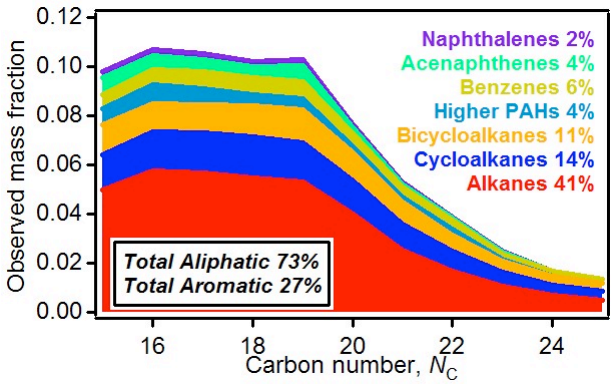


Figure 3. Composition of analyzed diesel fuel sample as fraction of total observed mass of each DBE class at each carbon number in the range of C_{15} to C_{25}. DBE classes shown are those described in Table 1, which account for 82% of total observed mass. Calibration is approximate and based on measured responses of hexadecane for N_{DBE} = 0 through 3, dodecylbenzene for N_{DBE} = 4 through 6, and phenanthrene for N_{DBE} = 7 through 11.

Though photon energy does not significantly affect fragmentation patterns, it is potentially useful in the identification of unknowns. Comparison of diesel fuel ionization at 10.5 eV and 10.0 eV supports the assumption that most alkanes in the sample are saturated and that N_{DBE} = 1 is likely comprised primarily of cycloalkanes, not alkenes. This information on saturation, coupled with N_{DBE} and N_C obtained by GCxMS analysis, provides critical constraints for elucidating the structure of an unknown compound.

3.3. Separation of structural isomers

GC-VUV-MS was used to separate and identify structural isomers that are unresolvable using GC-EI-MS. Alkanes are typically quantified by EI-MS using ions with m/Q = 57 Th (the $C_4H_9^+$ fragment), while VUV-MS allows quantitation based on parent ions; an example of this is shown in Figure 4, which compares VUV and EI chromatograms of diesel fuel in the C_{16} to C_{19} retention time window. The GC-EI-MS chromatogram of 57 Th clearly exhibits large *n*-alkane peaks, interspersed with smaller peaks representing branched isomers. The straight-chain alkanes can be identified by retention time matching to standards as well as the presence of parent ions, though these are two orders of magnitude smaller than the quantitation ion. However, the branched isomers are undetectable on their parent ions and therefore cannot be identified without authentic standards. Using GC-VUV-MS, each chromatographic peak can clearly be identified by parent ions, allowing identification and quantitation of branched isomers. In many cases, the branched regions observed on 57 Th using GC-EI-MS are shown by GC-VUV-MS to be composed of alkanes of different carbon numbers.

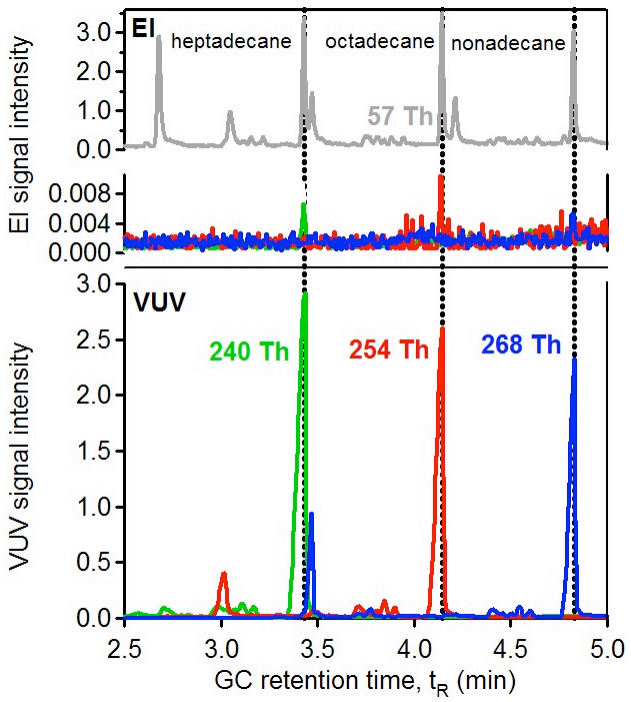


Figure 4. Chromatograms of diesel fuel using GC-EI-MS (top) and GC-VUV-MS (bottom). Ions shown are the parent ions of C_{17} (green line, 240 Th), C_{18} (red line, 254 Th), and C_{19} (blue line, 268 Th) saturated alkanes using both EI and VUV, and 57 Th (gray line) using EI, which is $C_4H_9^+$, the ion typically used for the quantitation and identification of alkanes. Dashed lines show *n*-alkanes. In EI, parent ions are visible for *n*-alkanes, but are below the level of quantitation in some cases, and isomers cannot be identified or resolved, whereas VUV allows for quantitation of straight-chain and branched isomers.

The GCxMS plot (Figure 2) shows several of structural isomers in each DBE class in diesel fuel. Alkanes (red circles in Figure 2) will be used to elucidate isomer separation as they are the largest fraction of diesel, have the clearest isomer patterns, and are not well-resolved using EI-MS. Isomers of a given N_C shift toward shorter retention times as branching increases, so saturated alkane isomers can be compared by normalizing retention times of isomers to the retention time of *n*-alkanes using a Kovat's index (I_K) transformation (Figure 5a).[25] The distance of each isomer from the *n*-alkane is given as the difference (ΔI_K) between the Kovat's index of the isomer and of the *n*-alkane of the same carbon number; by definition $\Delta I_K = 0$ for *n*-alkanes while the earliest-eluting peak with a given N_C has the highest ΔI_K value.

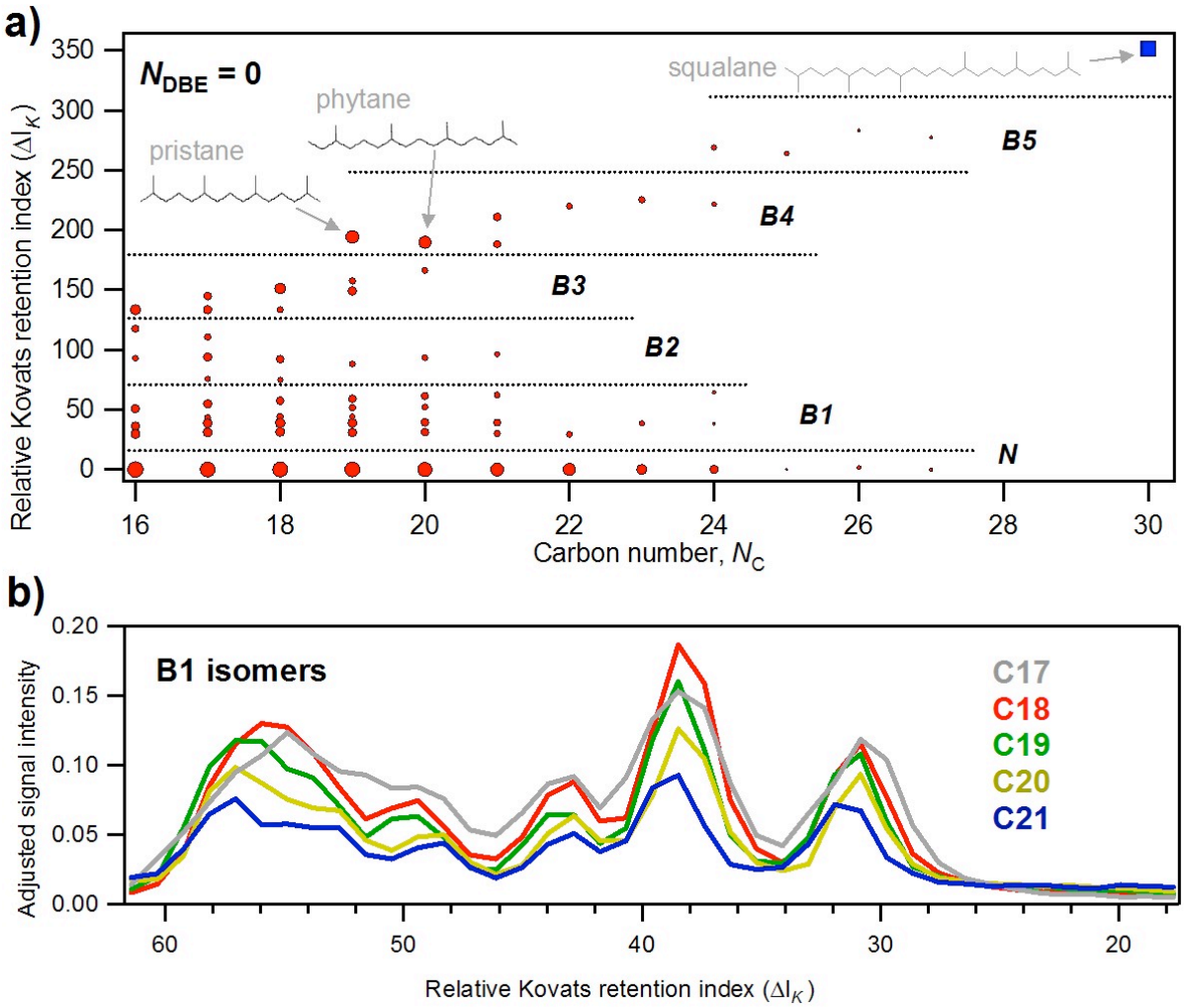


Figure 5. (a) Isomer separation of saturated alkanes demonstrated by relative Kovats retention index, measured as the difference between the Kovats index (I_K) of a given peak and that of the *n*-alkane (N) of the same carbon number ($\Delta I_K = 0$). Red circles are peaks found from GCxMS analysis of diesel fuel while the blue square represents squalane, a C_{30} hexamethyl saturated alkane (shown). Pristane and phytane, C_{19} and C_{20} tetramethyl saturated alkanes respectively (shown), are identified by authentic standards. Peaks fall into naturally separated regions shown by dashed lines through the gaps, defining areas of increasing branching from methyl (B1) through C_5-branching (B5). Squalane is in a B6 region that is unpopulated by alkanes observed in this diesel sample. Further discussion

of these characterizations can be found in Section 3.3. (b) Reproducibility of single methyl group isomer (B1) patterns of saturated alkanes ($N_{DBE} = 0$) from C_{17} through C_{21}. Detector response has been adjusted for GC transfer efficiency so all peaks are on equivalent scales, assuming approximately equivalent detector responses and absorption cross-sections.

Isomer structure can be determined from ΔI_K. The branched endmembers in the C_{19} and C_{20} saturated alkane bands are identified using authentic standards as tetramethyl isomers: pristane and phytane (structures shown in Figure 5a). These compounds are two of the most dominant alkanes observed in diesel fuel and are amongst only a few branched isomers regularly quantified. For C_{19} and C_{20}, the range of saturated alkane isomers is therefore bracketed by the *n*-alkane (N) and a C_4-branched isomer (B4), with less-branched isomers in between. Due to a lack of available standards, the dependence of retention time shifts on branching location cannot be determined, so an isomer with 4 methyl groups cannot be distinguished from an isomer with a single branch containing a butyl group; therefore this work will refer to C_n-branched (as B*n*) isomers in place of specific identification. Intermediately branched isomers appear to fall into approximately evenly spaced regions separated by natural gaps, suggesting C_1-, C_2-, and C_3-branching. Pristane has a ΔI_K significantly larger than the most branched C_{18} alkane, thus defining the boundary with the B3 region. A similar increase in ΔI_K of the most branched alkane occurs at C_{24}, defining a region of C_5-branched alkanes (B5). Squalane, a hexamethyl C_{30} alkane (blue square in Figure 5a), displays another significant increase in ΔI_K, suggesting a region of C_6-branched (B6) isomers that is unpopulated in the diesel sample analyzed in this work. Boundaries of the regions of C_1- through C_3-branching are inferred from gaps in peaks, though regions B2 and B3 do not have a clear separation point in this sample. Branched end members and *n*-alkanes are the largest peaks at most carbon numbers, so resolved B1 through B3 alkanes are below detection limits for larger alkanes, though some unresolved mass is often present.

As carbon number increases, the degree of possible branching increases correspondingly, so an increase in the width of the alkane band is expected. However, ΔI_K exhibits not a smooth upward trend, but rather a step function that is punctuated by increases of 30 to 50 every 5 carbon numbers, suggesting a theoretical maximum of branching, with a branch on every fifth carbon such as in the case of pristane. These compounds are thought to be polymers of isoprene (2-methyl-1,3-butadiene, C_5H_{10}), a biological carbon backbone commonly observed in plants,[26] biogenic emissions[27] and insects.[18] Only one B4 C_{19} isomer is observed, though the presence of ethyl or propyl branches would allow for multiple possible B4 C_{19} isomers while maintaining five carbons between branches. This suggests a preference for methyl groups and an absence of more complex branching patterns, in agreement with the conclusion that these higher-branched alkanes have an isoprenoid origin. These rules allow fewer isomers as branching increases, supported by the large number of B1 peaks at each carbon number compared to relatively fewer isomers of higher branching. In this diesel sample, an alkane of a given carbon number typically has 8-10 major isomers, despite the tens or even hundreds of thousands of isomers that are theoretically possible.[28]

These isomer regions are reproducible between carbon numbers, as demonstrated by Figure 5b, the comparison of C_{17} through C_{21} methyl isomers. Each carbon number consists of four well-resolved methyl isomers and a broad peak centered at $\Delta I_K \approx 56$ that is likely a co-elution of at least two isomers. The number of methyl isomers observed is a significant fraction of the number of possible isomers; a C_{18} alkane, for instance, has 8 possible B1 isomers, of which 5 or 6 are observed. The same isomers have similar relative intensities at each carbon number, assuming here that absorption cross-section

and detector response are similar for these compounds as would be expected. The similarity of the distribution of B1 isomers across all carbon numbers suggests a process either in formation or in petroleum refining that constrains isomer composition.

Quantification of isomers provides an estimate of the importance of branched alkanes in this diesel fuel sample (Figure 6). B4 isomers are quantified based on pristane and phytane, which have similar response factors but fragment approximately twice as heavily as *n*-alkanes. This fragmentation calibration factor is unknown for B1 through B3 isomers, but it is expected to fall between *n*-alkanes and B4 isomers, providing upper and lower bounds. Figure 6 shows that as saturated alkane carbon number increases, the total mass-fraction of branching increases from approximately 0.4 to approximately 0.6, which may be due to the increase in number of possible isomers. The B4 isomer is largest at the lowest carbon numbers for which it is present, a pattern that appears to apply to B3 and B5 isomers and may be related to being polymers of isoprene. Though B5 isomers are a large fraction of C_{24} through C_{26} alkanes, these low-volatility alkanes are a small fraction of total observed alkane mass. In total, branched isomers account for 26 to 44% of saturated alkane mass in this diesel fuel sample. Methyl isomers, which are difficult to resolve by GC-EI-MS and therefore not often quantified, are the largest fraction, comprising 10 to 16% of all alkane mass. B4 isomers are a significant fraction of saturated alkanes (5 to 9%), but quantifying only pristane, phytane, and a few other prevalent branched alkanes underestimates branching in diesel fuel by approximately half. Furthermore, understanding not just the mass of branched isomers, but also identifying and classifying them by carbon number extends the utility of this analysis into understanding compound-specific reactions and interactions.

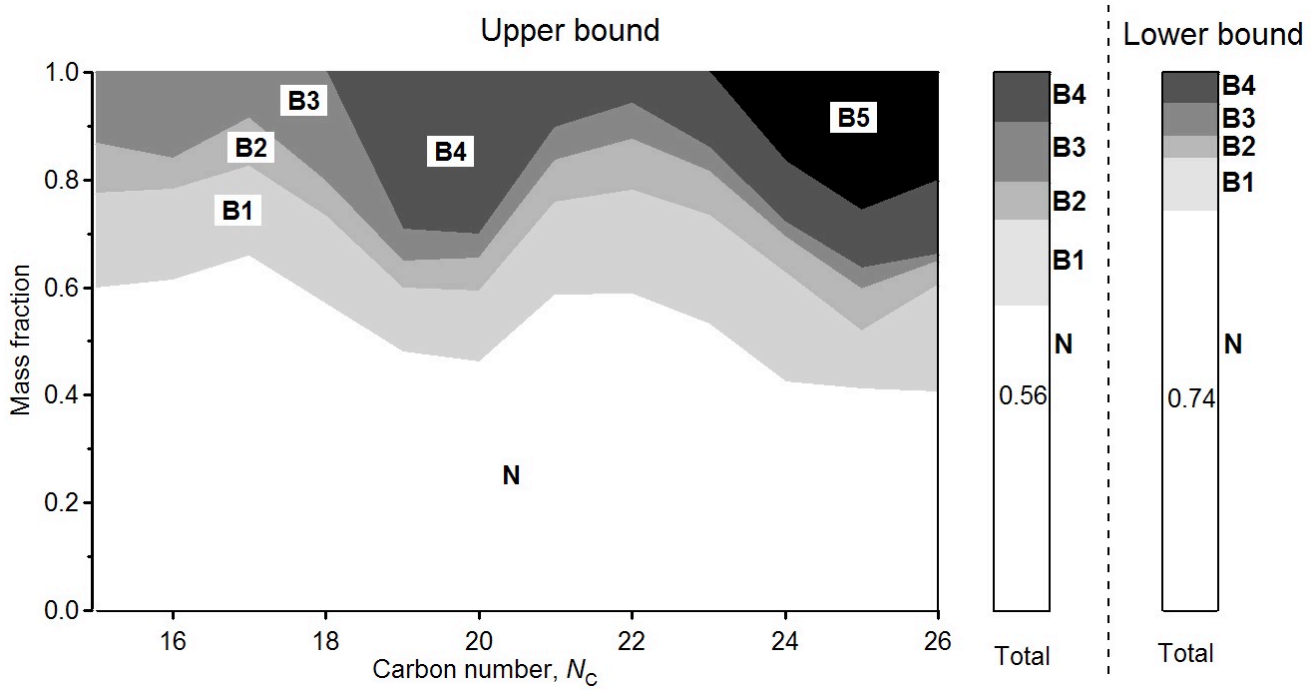


Figure 6. Mass fraction of each saturated alkane (N_{DBE} = 0) carbon number composed of branched isomers of type B1 (methyl), B2 (C_2-branched e.g. dimethyl),…B5 (C_5-branched e.g. pentamethyl). Branched isomers have been normalized to *n*-alkanes (N) by f_P. Upper bound calibration assumes f_P for B1 through B3 is similar to B4 and B5 (half that of the *n*-alkane), while lower bound calibration assumes f_P for B1 through B3 is similar to that of the *n*-alkane. Total fraction on right is the fraction at each carbon number weighted by the distribution of alkanes (Figure 3), showing branched isomers account for between 26 and 44% of total alkane mass.

Branched isomers of higher aliphatic DBE classes are uniquely resolvable using GC-VUV-MS. Cycloalkanes (N_{DBE} = 1) are shown in Figure 5a to be composed of 10 to 12 different isomers. Quantitative analysis shows these isomers to account for approximately half of the mass of cycloalkanes, though isomer distribution patterns are not as reproducible or clear as those of saturated alkanes. Analysis of complex mixtures by GC-EI-MS would therefore be likely to underestimate the importance of branching not only for saturated alkanes, but for other aliphatic hydrocarbons as well. Furthermore, GC-VUV-MS allows the branching of previously unknown compounds to be estimated from ΔI_K, further elucidating structure. The utility of the classification scheme developed in this work can be demonstrated on the unknown peak labeled "C" in Figure 2 that is unresolved and unidentified by GC-EI-MS and lies in the middle of the "unresolved complex mixture." From carbon number separation and DBE class, this compound is expected to be a C_{18} saturated cycloalkane, while its shift in t_R suggests that this compound has a single methyl branch; the structure of an otherwise unresolved compound can therefore determined.

4. CONCLUSIONS

Soft ionization coupled to chromatographic separation allows characterization of compounds based on structural characteristics and mass. GCxMS analysis of complex mixtures was used to classify unknown compounds by carbon number (N_C), number of rings and double bonds (N_{DBE}), and degree of branching based on retention time shifts (ΔI_K). Classification by these parameters provides structural information about unknown compounds in environmentally-relevant hydrocarbon mixtures, significantly improving

the resolution and identification of aliphatic hydrocarbons that cannot be differentiated using traditional MS techniques. The characterization scheme described in this work therefore provides significantly improved resolution of the "unresolved complex mixture" often present in GC analysis of environmental samples. The structure of any unknown compound in a complex mixture can be well-constrained by N_C, N_{DBE}, and ΔI_K, allowing estimation of reactivity, vapor pressure, and other physical properties that can be useful across a variety of scientific fields. Demonstration of these classification parameters on diesel fuel highlights the importance of isomer analysis. Quantitative analysis of branching using ΔI_K is useful in estimating the prevalence of branched isomers in environmental samples and is expected to provide significant insight into the formation and degradation of contaminants. Furthermore, the mass spectra of known compounds compiled in the Supplementary Information provides a useful reference for other GC-VUV-MS as well as non-chromatographic VUV-MS applications.

Improved resolution and isomer separation capabilities provided by GC-VUV-MS and GCxMS analysis are useful to any scientific fields that routinely characterize complex hydrocarbon mixtures, e.g. atmospheric chemistry, petroleum refining, microbial and chemical ecology, and bioremediation. Due to differences in chemical decomposition pathways of constitutional isomers, characterization of hydrocarbon branching has potential applications to the modeling of petroleum degradation in atmospheric and biological systems. Furthermore, some biological systems have been shown to be sensitive to isomer composition and branching: for instance, ants use long-chain methyl and dimethyl alkanes to determine colony-mates.[19,20] The ability to quantitatively separate isomers and constrain the structure of unknown compounds in

complex hydrocarbon mixtures will be useful in the analysis of a variety of samples from wide-ranging disciplines.

Supporting Information Available: This material is available free of charge via the Internet at http://pubs.acs.org.

ACKNOWLEDGEMENT This Advanced Light Source as well as K.R.W. and T.N. are supported by the Director, Office of Energy Research, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. This work was also supported by the Laboratory Directed Research and Development Program of Lawrence Berkeley National Laboratory under U.S. Department of Energy Contract No. DE-AC02-05CH11231, the US Department of Energy STTR Project DE-FG02-08ER-86335, and the National Science Foundation Grant CHE-1012809 and Atmospheric Chemistry Program Grant 0931934. G.I. was funded by the U.S. Environmental Protection Agency (EPA) Science to Achieve Results (STAR) program, Fellowship Assistance Agreement no. FP-91781901-0. This work has not been formally reviewed by EPA. The views expressed in this work are solely those of the authors, and EPA does not endorse any products or commercial services mentioned.

REFERENCES

(1) Megharaj, M.; Ramakrishnan, B.; Venkateswarlu, K.; Sethunathan, N.; Naidu, R. *Environ. Int.* **2011**, *37*, 1362-1375.

(2) Simoneit, B. R. *Mass Spec. Reviews* **2005**, *24*, 719-765.

(3) Fraser, M. P.; Lakshmanan, K. *Environ. Sci. Tech.* **2000**, *34*, 4560-4564.

(4) Mazurek, M. A. *Environ. Health Perspect.* **2002**, *110*, 995-1003.

(5) Schauer, J. J.; Rogge, W. F.; Hildemann, L. M.; Mazurek, M. A.; Cass, G. R. *Atmos. Environ.* **1996**, *30*, 3837-3855.

(6) Williams, B. J.; Goldstein, A. H.; Kreisberg, N. M.; Hering, S. V.; Worsnop, D. R.; Ulbrich, I. M.; Docherty, K. S.; Jimenez, J. L. *Atmos. Chem. Phys.* **2010**, *10*, 11577-11603.

(7) Eschner, M. S.; Welthagen, W.; Gröger, T. M.; Gonin, M.; Fuhrer, K.; Zimmermann, R. *Anal. Bioanal. Chem.* **2010**, *398*, 1435-1445.

(8) Hanley, L.; Zimmermann, R. *Anal. Chem.* **2009**, 4174-4182.

(9) Sieck, L. W. *Anal. Chem.* **1983**, *55*, 38-41.

(10) Adam, T.; Zimmermann, R. *Anal. Bioanal. Chem.* **2007**, *389*, 1941-1951.

(11) Zimmermann, R.; Mühlberger, F.; Fuhrer, K.; Gonin, M.; Welthagen, W. *J. Mater. Cycles Waste Manag.* **2008**, *10*, 24-31.

(12) Mysak, E. R.; Wilson, K. R.; Jimenez-Cruz, M.; Ahmed, M.; Baer, T. *Anal. Chem.* **2005**, *77*, 5953-5960.

(13) Geissler, R.; Saraji-bozorgzad, M. R.; Gro, T.; Fendt, A.; Streibel, T.; Sklorz, M.; Krooss, B. M.; Fuhrer, K.; Gonin, M.; Kaisersberger, E.; Denner, T.; Zimmermann, R. *Anal. Chem.* **2009**, *81*, 6038-6048.

(14) Mitschke, S.; Welthagen, W.; Zimmermann, R. *Anal. Chem.* **2006**, *78*, 6364-6375.

(15) Schauer, J. J.; Kleeman, M. J.; Cass, G. R.; Simoneit, B. R. *Environ. Sci. Tech.* **1999**, *33*, 1578-1587.

(16) Kroll, J. H.; Seinfeld, J. H. *Atmos. Environ.* **2008**, *42*, 3593-3624.

(17) Lim, Y. B.; Ziemann, P. J. *Environ. Sci. Tech.* **2009**, *43*, 2328-2334.

(18) Blomquist, G. J., Bagneres, A., Eds. *Insect Hydrocarbons*; Cambridge University Press: Cambridge, UK, **2010**.

(19) Martin, S. J.; Helanterä, H.; Drijfhout, F. *P. Proc. R. Soc. B.* **2011**, *278*, 496-503.

(20) van Wilgenburg, E.; Sulc, R.; Shea, K. J.; Tsutsui, N. D. *J. Chem. Ecol.* **2010**, *36*, 751-788.

(21) Youngsteadt, E.; Guerra Bustios, P.; Schal, C. *PloS One.* **2010**, *5*, e15822.

(22) Gloaguen, E.; Mysak, E.; Leone, S.; Ahmed, M.; Wilson, K. *Int. J. Mass Spectrom.* **2006**, *258*, 74-85.

(23) Heimann, P. A.; Koike, M.; Hsu, C. W.; Blank, D.; Yang, X. M.; Suits, A. G.; Lee,

(24) NIST Mass Spec Data Center, S.E. Stein, director. Mass Spectra. In *NIST Chemistry WebBook, NIST Standard Reference Database Number 69*; Linstrom, P. J., Mallard, W. G., Eds.; National Institute of Standards and Technology: Gaithersburg, MD, **2011**.

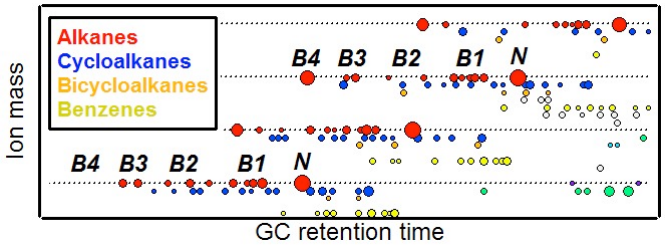
(25) Kovats, E. *Helv. Chim. Acta*. **1958**, *41*, 1915-1932.

(26) Langenheim, J. H. *J. Chem. Ecol*. **1994**, *20*, 1223-1280.

(27) Guenther, A.; Hewitt, C. N.; Erickson, D.; Fall, R.; Geron, C.; Graedel, T.; Harley, P.; Klinger, L.; Lerdau, M.; Mckay, W. A.; Pierce, T.; Scholes, B.; Steinbrecher, R.; Tallamraju, R.; Taylor, J.; Zimmerman, P. *J. Geophys. Res*. **1995**, *100*, 8873-8892.

(28) Goldstein, A. H.; Galbally, I. *Environ. Sci. Tech*. **2007**, *41*, 1514-1521.

for TOC only


Alkanes
Cycloalkanes
Bicycloalkanes
Benzenes
B4 B3 B2 B1 N
B4 B3 B2 B1 N
Ion mass
GC retention time

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor the Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or the Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or the Regents of the University of California.