ADVANCING REGULATORY DECISION MAKING OF UVCB SUBSTANCE THROUGH NOVEL ANALYTICAL TECHNIQUES

A Dissertation

by

ALINA TERESA ROMAN-HUBERS

Submitted to the Graduate and Professional School of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Ivan Rusyn
Committee Members,	Weihsueh Chiu
	Thomas J. McDonald
	Delina Lyon
	Natalie Johnson
Head of Department,	Ivan Rusyn

May 2022

Major Subject: Toxicology

ABSTRACT

Substances of unknown and variable composition, complex reaction product and biological material (UVCB) are one of the most challenging areas in regulatory science and risk assessment. The obscured compositional characterization and understanding of these substances, such as petroleum products, have relied in broad physical-chemical properties to deduce their toxicological fate and behavior. This study aims to advance the risk assessment of UVCB substances through high-resolution mass spectrometry to comprehensively deconvolute their inherently complex and variable chemical composition at a molecular level. Ion mobility coupled with quadruple-time of flight- mass spectrometer (IMS-MS) was employed to rapidly screen petroleum UVBC substances to define the value of petroleomic characterization in regulatory decision making. First, we employed comprehensive IMS-MS acquired compositional profiles of crude oil for categorization based on their geographical source. We compared high-resolution information with that from conventional methods, establishing IMS-MS as a high-throughput and highly informative tool for the evaluation of UVCB substances. Then, we demonstrated the value of isobaric and isomeric characterization afforded by IMS-MS to comprehensively define molecular compositional profiles of petroleum refined products. High-resolution qualitative and semi-quantitative compositional information served to measure the degree of commonality across production cycles and between categories of UVCB substances to assist read-across and forensic fingerprinting assessments. Lastly, we assessed the environmental fate and behavior of weathered and photooxidized oil slick

with high-resolution characterization of the dissolved organic matter from a mesocosm experiment. Through evaluation of molecular compositional trends, we determined that the physicochemical transformations of the composition of UVCBs correlated with sunlight irradiation achieving a better understanding of the environmental fate of an oil slick due to weathering to define the potential exposures and hazards. Together, this work demonstrated the utility of IMS-MS for rapid and comprehensive structural and molecular characterization of petroleum UVCB substances at a qualitative and semi-quantitative level. Through this dissertation we contribute a novel approach for the characterization of UVCB substances ensuring no underestimation of related human health and environmental hazards to support regulatory science and safeguard the risk assessments.

DEDICATION

I would like to dedicate this dissertation to my parents, Dr. Radames Roman Grau and Hendriekje Hubers, and my brother, Alvaro Roman-Hubers. Without their unconditional support and guidance, I would not be where I am today. I would also like to dedicate this dissertation to the Interdisciplinary Faculty of Toxicology at Texas A&M University, especially my mentor Dr. Ivan Rusyn. I will forever be grateful for the opportunities you have offered me to grow into the person and scientist I am, thank you for your guidance and support.

ACKNOWLEDGEMENTS

I would like acknowledge and give the warmest thanks to my committee chair, Dr. Ivan Rusyn, for his unwavering guidance and mentorship throughout my graduate career. I am grateful for the resources and opportunities Dr. Rusyn has offered me to continuously grow and to sharpen my skills, work ethics and critical thinking. Additionally, I am would like to thank Dr. Weihsueh Chiu, Dr. Thomas McDonald, Dr. Delina Lyon and Dr. Natalie Johnson for serving in my doctoral dissertation committee. I am very grateful to my committee members for believing in my potential and offering the opportunity of transferring from the master's program to pursue my doctorates degree.

I would like to express my gratitude to the current and past members of the Rusyn Lab for their support and friendship. Special thanks to Dr. Zunwei Chen, Dr. Yizhong Liu, Dr. Courtney Sakolish and Lucie Ford for their patient and continuous support in my cell culture training. Thanks also to Dr. Yu-Syuan Luo, Dr. Noor Aly, Alan Valdiviezo, Alexandra Cordova and the Baker Lab, Dr. Erin Baker and Dr. James Dodds, for their analytical chemistry and IMS-MS technical support. Special thanks to the Dr. Fred Wright and Dillan Lloyd for their contribution in the statistical analysis of large data sets. I would also like to thank our collaborators, contributors and funding sources without whom this work would not have been possible. Thanks to Dr. Tom Parkerton, Dr. Kelly McFarlin and Dr. Christoph Aeppli for including us in their ongoing mesocosm exposure study. Special thanks to the Geochemical Environmental Research Group at Texas A&M University for facilitating us the crude oil samples used in this research.

I am very grateful to the Interdisciplinary Faculty of Toxicology, specifically Kim Daniel, for their support and guidance throughout the ups and downs of graduate school, as well as for offering us networking and extra-curricular training opportunities. I would like to recognize the training offered to me by the Texas A&M Superfund Research Center and Arlean Rohde for all her hard work and dedication to the program and trainees.

I would like to acknowledge my colleagues from my internship at Chevron for their guidance and collaboration. I would like to particularly single out my supervisors and mentors, Dr. Denise Roesh and Dr. Ziad Naufal. I would like to thank you for your support and for all the opportunities I was given to further my education and professional development.

Lastly, I would like to acknowledge those who have been my support group since the day I got my acceptance letter my parents (Radames and Hendriekje), my brother (Alvaro), my extended family (Cuca, Chiqui, Pancho, Randy, Mercedes, Waleska, Mauricio, Isabella and Ignacio). Thank you for supporting my dream, encouraging me to work and for always letting me know you were by my side. I am very grateful to my colleagues who have become dear friends Zunwei Chen, Alan Valdiviezo, Lucie Ford, Haley Moyer, Courtney Sakolish and Alicia Lim who have who have joined me throughout this journey. Special thanks to my friends (Natalia, Patricia, Luis, Angeli, Ben, Fravelyn and Luara Lynn) for always being there just a phone call away. I would also like to thank my grandmothers, Adela Hubers and Teresa Roman, for being my guardian angels, teaching me the value of working hard and for paving the way for me to be achieving my dreams.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Professor Ivan Rusyn (advisor), Weihsueh Chiu, Thomas J. McDonald, Delina Lyon and Natalie Johnson.

Statistical analysis assays in this dissertation were performed with the technical guidance from Dr. Weihsueh Chiu at Texas A&M University and Dr. Fred Wright and Dillon Lloyd from North Carolina State University. Technical support about IMS-MS was provided by Dr. Erin Baker and Dr. James Dodds from North Carolina State University. Technical support and access to GC-MS instrumentation was provided by Dr. Thomas J. McDonald at Texas A&M University. The crude oil sampled analyzed in Chapter 2 were kindly provided by the Geochemical Environmental Research Group at Texas A&M. The refined oil samples evaluated in Chapter 3 were contributed by Arlean Rhode. The water samples studied for Chapter 4 were collected and provided by ExxonMobil and Bigelow Lab.

All other work conducted for the thesis (or) dissertation was completed by the student independently.

Funding Sources

This study was supported, in part, by grants from the National Institutes of Health (P30 ES029067 and P42 ES027704) and the National Academies Gulf Research Program (2000008942). A.T. Roman-Hubers was supported, in part, by a training grant from the National Institutes of Health (T32 ES0226568).

NOMENCLATURE

UVCB	Unknown or Variable composition, Complex reaction products, or Biological materials
REACH	Registration, evaluation, authorization and restriction of chemicals
CAS	Chemical Abstract Service
ASTM	American Society of Testing and Materials
MS	Mass Spectrometry
FTICR-MS	Fourier-Transform Ion Cyclotron Resonance Mass Spectrometer
Orbitrap-MS	Orbitrap Mass Spectrometer
TOF-MS	Time of Flight Mass Spectrometer
IMS-MS	Ion Mobility Mass Spectrometer
GC-MS	Gas Chromatography Mass Spectrometer
m/z	Mass-to-Charge
$^{DT}CCS_{N2}$	Nitrogen gas-filled drift tube collisional cross section
ESI	Electrospray Ionization
APPI	Atmospheric Pressure Photoionization
APCI	Atmospheric Pressure Chemical Ionization
FM	Fowlkes-Malow
PAHs	Polycyclic Aromatic Hydrocarbons
KMD	Kendrick Mass Defect
DBE	Double Bond Equivalent

FWHM	Full Width at Half Maximum

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

1. INTRODUCTION

1.1. Overview

Crude oils are naturally occurring complex and variable substances composed of thousands of constituents and hydrocarbon molecules. Billions of barrels of crude oils are processed yearly yielding a wide variety of refined petroleum substances (Kaiser 2017; Salvito et al. 2020). The intrinsic composition of petroleum substances is defined by the crude oil from which it is derived and the manufacturing and refining processes it undergoes (i.e. distillation streams, boiling point). The inherent compositional complexity and variability of petroleum substances that define these products classify them as substances of unknown, variable composition, complex reaction products, or biological materials (UVCBs). The indefinite and unpredictable molecular composition of UVCBs presents unique challenges for substance assessment and regulation (Clark et al. 2013; ECHA 2017a). The broad range in compositional properties of petroleum UVCBs challenges the standard testing paradigm designed for single substances. Therefore, hazard, risk and exposure assessment of petroleum substances have heavily relied on substance grouping and read-across. Additionally, new approach methods such as in silico and in vitro methods, have been proposed in the assessment of UVCBs.

The complex composition of petroleum substances influences the fate and concentration of each of its constituents. Therefore, a comprehensive understanding of the chemical composition of petroleum substances is needed to be enable appropriate risk assessment.

1.2. Regulatory framework for assessment of UVCB substances

International regulatory frameworks have historically (Figure 1.1) attempted to address petroleum UVCB substance registration through supplementary information, such as physical-chemical properties, distillation/manufacturing processes, and chemical fingerprint (Dimitrov et al. 2015; ECHA 2008; Rasmussen et al. 1998). Naming conventions, developed by American Petroleum Institute and US Environmental Protection Agency to be included in the Toxic Substance Control Act inventory, were based on the substances manufacturing processes (API 1983; EPA 1995; U.S. EPA 1978). Substance categories were first defined through the High Production Volume Challenge Program with the compiling of substance information, physical-chemical properties, and human and environmental hazards, voluntarily provided by industry (Group 2017). The European Union Regulation No 1907/2006 Registration, Evaluation, Authorization, and Restriction of Chemicals (REACH) was implemented in 2007, shifting responsibility to industry partners in the safeguarding of assessing the potential risk posed by chemicals. By 2010, all petroleum products, 8,000 individual registrations, were registered in the Europe under REACH deadline for substances produced at >100 tones (Concawe 2022). Of the 191 substances registered a limited amount have complete hazard assessment data sets (Concawe 2021).

Based on the available guidance and intrinsic composition of petroleum substances, industry has relied on categorization of substances through extrapolation of available information to fill in data gaps without the need for further testing (Clark et al. 2013; Pusyn et al. 2009). Categorization of substances has been defined through broad

compositional commonalities and predicting a correlation to similar physiochemical and toxicological properties (Clark et al. 2013; McKee et al. 2015; OECD 2014). Representative substances known as "worst-case scenario" are used to read-across data for substances in the cohort, ensuring no underestimation and focusing on constituents that are known to drive the hazard potential of the substance (Clark et al. 2013; McKee et al. 2015; Salvito et al. 2020). Additionally, recent amendments to Annex XI, S.1.5 of the REACH regulation state that for the application of grouping "*Structural similarity for UVCB substances shall be established on the basis of similarities in the structures of the constituents, together with the concentration of these constituents and variability in the concentration of these constituents*" (European Commission 2021).

1.3. Standardized methods for the characterization of UVCB substances

The developed and validated methods for characterization of UVCB substances enable appropriate substance description for determining product quality and assessing potential human health and environmental hazards. Industry partners follow standardized methods in the characterization of substances for registration to satisfy the requirements established by regulatory agencies. Information for substance registration following REACH should be sufficient to enable substance identification. Therefore, European Chemical Agency's guidance is that analytical data of a substance for registration should provide: source/feedstock; refining history; boiling and carbon range; identification and concentration of constituents present at $\geq$10%, of relevant hazard classification or/and PBT assessment; identification of additives; identification of unknown constituents through generic description; chromatographic or spectral information; flash point and

viscosity (CONCAWE 2012; ECHA 2017a). Additionally, Article 7(2) and 33 of REACH have defined a concentration threshold of 0.1% w/w for constituents classified as "substance of very high concern" to be notified or communicated in safety data sheets (ECHA 2017b).

Due to the broad range of UVCB substances with widely different compositions, volatilities and polarities it is not possible to define a one size fits all technique. Therefore, characterization of UVCB substances has been an active research area in analytical chemistry and the continuous advances in technology and methods have contributed to a more detailed understanding of these highly complex substances (Onel et al. 2019; Stout and Wang 2007; Wang et al. 2011). The analytical information for REACH registration dossiers has been defined to that achievable by standardized methods (CONCAWE 2014). Accordingly, industry partners have relied on a battery of standardized tests to characterize substances. The first established methods characterized petroleum substances by their specific gravity, *ASTM D287 Standard Method for API Gravity of Crude Petroleum and Petroleum Products* (Giles 2016). Further physical-chemical information such as boiling and carbon range description can be deduced through physical distillation (ASTM D86, D1160 and D2892) or simulated distillation methods (CONCAWE 2019a; 2020a). Likewise, elemental analysis has been fundamentally applied for the characterization of petroleum UVCBs measuring the concentrations of major elements ranging from organics to metals (CONCAWE 2019a).

Spectroscopic techniques have been widely employed to obtain broad compositional information for regulatory characterization and identification of UVCB

substances; nevertheless the value of this information has been challenged (CONCAWE 2020a). Ultra-violet spectroscopy analysis quantifies compounds by detecting unsaturated bonds such as those in olefins and aromatic bonds, ketonic and heteroatom groups. Through infra-red spectroscopy measurements the presence of functional groups can be determined allowing one to define the degree of saturation in the constituents (CONCAWE 2012). Nuclear magnetic resonance methods (IP392, ASTM D5292) measure the percent of carbon or hydrogen atoms in an aromatic ring (CONCAWE 2020a).

Chromatographic based methods have been used to separate constituents providing information on the total content of those present. Liquid chromatography separation is dependent on the polarity of the constituents present, predominantly used to characterize polar compounds. High-performance liquid chromatography is used to quantify mono-, di- and tri-aromatic hydrocarbons. Meanwhile, thin layer or liquid column chromatography (ASTM D2007) separation generates information on basic chemical functionalities present (CONCAWE 2012; 2019a). Gas chromatography analysis of hydrocarbons was first applied by Eggerston in 1960, and later published by Green in 1964, leading to the standardized method ASTM 2887-84 for *Boiling Range Distribution of Petroleum Fractions by Gas Chromatography* (Giles 2016). Gas chromatography has been a powerful tool used for the separation and semi-qualitative assessment of non-polar constituents such as hydrocarbons and polycyclic aromatic hydrocarbons (PAHs) (CONCAWE 2012). Gas chromatography-mass spectrometry (GC-MS) detection based methods have been heavily applied to forensic fingerprinting methods (US EPA 8270 and

8051B) to characterize the composition of petroleum UVCBs (US EPA 1996; 2014). Flame ionization detection has been coupled with various liquid and gas chromatography methods for the detection and quantification of hydrocarbons (CONCAWE 2012). Otherwise coupled with MS providing molecular structure information, the efficient separation of liquid and gas chromatography methods facilitate qualitative and quantitative characterization of the substances measured. The information generated through gas chromatography with mass spectrometry for quantitative characterization has demonstrated it as a valuable tool throughout production, regulatory assessment, and emergency response of petroleum substances.

1.4. Novel analytical methods for the characterization of UVCB substances

Advancements in mass spectrometry technology in the past 80 years (Figure 1.1) have spawned the application of high-resolution mass spectrometry for the study of petroleum substances at a molecular level, a field defined as petroleomics. Mass resolution has been defined as the minimum mass difference between two spectral peaks ($m_2 - m_1 = \Delta m_{50\%}$). The performance of a mass analyzer is usually expressed based on its resolving power ($m_2/ m_2 - m_1$). The resolving power of high-resolution mass spectrometry instruments, measured by full width at half maximum (FWHM), typically greater than 10,000 FWHM results in increased mass accuracy allowing one to differentiate closely spaced peaks. The high resolution and increased mass accuracy ensure peak resolution for molecular formula assignment to observed masses. The comprehensive characterization of the molecular composition of petroleum has defined the field known as “petroleomics” yielding information to predict the behavior and reactivity of petroleum substances (Hsu

et al. 2011; Marshall and Rodgers 2004; Niyonsaba et al. 2019; Palacio Lozano et al. 2019a; Palacio Lozano et al. 2019b; Palacio Lozano et al. 2020; Xian et al. 2012).

The Fourier-transform ion cyclotron resonance mass spectrometer (FTICR-MS) has an ultra-high resolving power of ~10^6 FWHM offering state-of-the-art mass resolution and accuracy. Orbitrap-MS, another Fourier-transform MS instrument offers a resolving power of ~10^5 FWHM. Recent advancements in time of flight mass spectrometer (TOF-MS) technology render a resolving power of 5^4-10^4 FWHM. Unlike FTICR-MS and Orbitrap-MS which offer ultra-high-resolution, TOF-MS rapid acquisition time allows for coupling with various separation techniques such as two-dimensional gas chromatography and ion mobility spectrometry (Palacio Lozano et al. 2019a). The high-resolution and mass accuracy offered by FTICR-MS, Orbitrap-MS, and TOF-MS, can completely separate coeluting mass spectral peaks present in complex substances allowing to calculate the exact molecular mass for the mass spectrum peak detected (Palacio Lozano et al. 2019b; Rodgers and McKenna 2011).

Petroleomic analysis afforded by high-resolution mass spectrometry has proven to be a powerful approach in the evaluating UVCB substances through its application in over 1,000 studies (Figure 1.2). The expanded window of analysis available by high-resolution mass spectrometry makes it possible to characterize less volatile and non-volatile, as well as polar and non-polar compounds within a broad mass range ($\geq$50 *m/z*) (Staš et al. 2017). High-resolution mass spectrometry technique's resolving power and mass accuracy can resolve closely spaced mass peaks in complex substances, resulting in useful molecular level characterization. The ability to detect a compound's exact molecular mass allows to

determine its molecular composition ($C_c H_h N_n O_o S_s$) (Staš et al. 2017). The application of Kendrick Mass Defect (KMD) has aided the molecular characterization through the conversion of the observed mass to the proposed Kendrick mass (Kendrick 1963). As a result, compounds of the same base units with different degree of alkylation (CH_2) would differ by 14.0000Da and would have the same defect in Kendrick Mass, making it possible to recognize homologous series of compounds of the same class and aromaticity (Figure 1.3A) (Hughey et al. 2001; Marshall and Rodgers 2004; 2008; Palacio Lozano et al. 2020).

1.5. Challenges in the regulatory assessment of UVCB substances

International regulatory frameworks have addressed the assessment of the potential hazards and risk related to petroleum substances based on their broad physical-chemical properties and manufacturing descriptors (Salvito et al. 2020). Regulators and industry partners have heavily relied on substance categorization and read-across assessment to extrapolate information for predicting a correlation to a substance with similar physiochemical and toxicological properties (Clark et al. 2013; McKee et al. 2015; OECD 2014; Pusyn et al. 2009; Salvito et al. 2020). Nevertheless, the chemical information has been limited to that requested by regulators, with broad compositional information for commercial purposes and no comprehensive information on individual constituents (ECHA 2020).

Conventional standardized methods recommended by regulatory agencies and industry partners for compositional characterization of UVCB substances are known to be impaired by the complex composition of UVCB substances. Therefore, it is currently necessary to carry-out a battery of assays, large sample quantities, specialized sample

preparation and considerable expenditures to acquire sufficient information for substance characterization. Despite the multiple available standardized methods and recent advancements in analytical technology, conventional analytical practice is still limited in the mass resolving power and struggles to separate the coeluting and isomeric compounds. Spectroscopic techniques are limited to bulk information and "*most substances* [petroleum UVCBs] *cannot be effectively differentiated from each other by UV, IR, ^{1}H-NMR or ^{13}C-NMR spectroscopies*" (CONCAWE 2020a). Chromatographic base methods provide sufficient information for qualitative and quantitative characterization of compounds within UVCB substances. Characterization of nonpolar and relative volatile compounds present in complex substances has readily relied on GC-methods such as ASTM D2134, D6729, D6730, among others (CONCAWE 2012). GC coupled with flame ionization detection and MS provide information on the aliphatic and aromatic fraction of complex substances readily applied for substance identification (Reddy and Quinn 1999; Wang and Fingas 2003b). Nevertheless, due to the limited resolution, compounds in heavy petroleum UVCBs are known to co-elute resulting in a hump defined as unresolved complex mixture (Wang et al. 2011; Weng et al. 2015). Characterization of constituents in the substances is also limited to prior selected compounds for analysis (i.e. biological markers) (Fernandez-Lima et al. 2009).

Risk assessment of UVCB substances has been under-characterized due to the substance's intrinsic composition and reliability on broad substance measurements. The European Chemical Agency in 2020 challenged the assumptions made by industry through the application of read-across for new substance registration, questioning the

uncharacterized "*broadly similar [chemical] composition*" and raising the need for additional chemical compositional descriptors to justify the correlation with currently available data. The European Chemical Agency stated the need of "*qualitative and quantitative information on composition to allow assessment of whether predictions are compromised*" (ECHA 2020). At the time, Annex XI, S.1.5 of the REACH regulation established that substances with chemical similarity driving their respective physicochemical, toxicological and ecotoxicology may be defined as a category (ECHA 2020). The amendments published in 2021 define the application of read-across to be justified by defined qualitative and quantitative structural similarity of constituents between substances resulting in the likelihood of similar toxicological properties (European Commission 2021).

1.6. Addressing regulatory challenges with high-resolution mass spectrometry

High-resolution mass spectrometry has provided detailed compositional characterization on petroleum UVCB substances (Hsu and Shi 2013; Niyonsaba et al. 2019; Palacio Lozano et al. 2019a; Palacio Lozano et al. 2019b; Palacio Lozano et al. 2020; Rodgers and McKenna 2011; Xian et al. 2012). The information available through high-resolution characterization of the chemical composition of UVCB substances can address the shortcomings with regards to the prediction of toxicological properties when practicing read-across assessment. Comprehensive characterization can provide sufficient information on the molecular composition and their relative abundance within a substance to define the commonality between substances and their similarity at a molecular level.

With this information, it is then possible to determine key constituents that have the potential of driving the human health hazard properties of the substance.

High-resolution mass spectrometry accompanied with petroleomic analysis of UVCB substances has generated information on "*individual chemical constituents, their structural features and quantitative variation*" addressing the challenges presented by ECHA (Figure 1.2) (ECHA 2020). The mass accuracy and resolving power afforded by FTICR-MS, Orbitrap-MS, and TOF-MS instrumentations have provided comprehensive isobaric information of the constituents in UVCB petroleum substances with high-resolution and mass accuracy (Niyonsaba et al. 2019; Palacio Lozano et al. 2019a; Palacio Lozano et al. 2019b; Palacio Lozano et al. 2020; Staš et al. 2017).The ultra-high-resolution analysis offered by FTICR-MS and Orbitrap-MS delivers an expanded mass-to-charge region, accurate mass measurements, and relative ion abundances allowing to visualize the compositional characteristics of light to heavy fractions of petroleum UVCBs (Kim et al. 2015; Miles et al. 2020; Palacio Lozano et al. 2020). The mass accuracy afforded by having ultra-high-resolution generates isobaric information of the constituents in a complex substance. Previous studies have benefited from the ability to distinguish compounds with the same nominal mass but different elemental compositions to comprehensively characterize the molecular formula of >95% of the compounds detected(Walters et al. 2015). Untargeted high-resolution mass spectrometry analysis has provided insight on specific compounds of interest to assess the compositional traits and behaviors of various substances (Cho et al. 2014; Walters et al. 2015).

In contrast, to other high-resolution mass spectrometers TOF-MS offers lower resolution, but a faster analysis time in the millisecond time scale. Therefore, facilitating the combination with various separation techniques such as two-dimensional gas chromatography and ion mobility spectrometry enhancing peak capacity and separation power (Palacio Lozano et al. 2019a). Ion mobility provides multidimensional separation of ions based on their mobility across a stationary buffer gas under the influence of an electric field (Dodds and Baker 2019). Mobility measurements are converted to collisional cross section values defining the spatial conformation of the ions and enabling isomeric measurement. Ion mobility spectrometry multidimensional separation with TOF-MS has been applied for the identification and characterization of structural isomers in complex substances (Hoskins et al. 2011; Lalli et al. 2015; Lalli et al. 2017; Mahmoud and Dabek-Zlotorzynska 2018). In comparison, two-dimensional gas chromatography orthogonal separation offers qualitative and semi-quantitative characterization of compound classes (Damasceno et al. 2014; O'Reilly et al. 2019; Stewart et al. 2021). The secondary retention time offers narrow separation resolving isomeric compounds and compounds that would otherwise coelute (Ball and Aluwihare 2014; Luna et al. 2014; Ngo et al. 2012; Rowland et al. 2011; Scarlett et al. 2008). The multidimensional gas chromatography separation coupled with high-resolution detection in TOF-MS has been widely employed to map out the composition of complex substances (Alam et al. 2018; Frenzel et al. 2010; Gabetti et al. 2021; Luna et al. 2014; Muller et al. 2020; O'Reilly et al. 2019; Qian and Wang 2019; Ristic et al. 2018; Zhu et al. 2020). Notably, physical characteristics (i.e. density) have

been directly correlated based on the detailed chemical composition defined through two-dimensional gas chromatography high-resolution mass spectrometry (Vozka et al. 2019).

As post-ionization separation techniques, high-resolution mass spectrometry instruments offer detection of non-polar and polar compounds through different ionization sources. The application of electrospray ionization (ESI) and atmospheric pressure photoionization (APPI) sources have expanded the characterization of polar and non-polar compounds, respectively (Oldenburg et al. 2014; Oldenburg et al. 2017; Walters et al. 2015). ESI has a higher selectivity to ionize polar compounds such as heteroatom containing constituents. On the other hand, APPI and atmospheric pressure chemical ionization (APCI) have been employed for the characterization of non-polar compounds, such as hydrocarbons and sulfur-containing compounds (Dong et al. 2019).

The comprehensive molecular characterization made possible by high-resolution mass spectrometry instruments and petroleomics analysis yields elemental and structural information (Figure 1.3). The broad information can be readily summarized to show the relative distribution of compounds through heteroatom classes ($HC_{\#}$, $O_{\#}$, $N_{\#,}$ $O_{\#}$, $S_{\#}$, etc.). It is also possible to better understand the structural composition and degree of aromaticity of the constituents by plotting DBE (DBE= $C_{\#}$ - $H_{\#}/2$ + $N_{\#}/2$ + 1) versus carbon number (Figure 1.3B) (Marshall and Rodgers 2008; Palacio Lozano et al. 2020). Van Krevelen diagrams allow one to define the degree of oxidation of the constituents by plotting the H/C versus O/C ratio of the organic compounds (Kim et al. 2003; van Krevelen 1950; 1984). Structural data such as that offered by two-dimensional gas chromatography or ion mobility spectrometry can complement the molecular information for compositional

categorization by hydrocarbon block method through grouping structurally related compounds by their carbon chain length and hydrocarbon class (Figure 1.3C) (CONCAWE 1996; 2019b).

The expanded analytical window offered by high-resolution mass spectrometry has improved the current understanding of the chemical profile of UVCB substances. The feasibility to readily evaluate heavy fuel oils (Elbaz et al. 2015) and trace molecular compositional changes of the heteroatomic compounds (Silva et al. 2021) has not been feasible until the application of high-resolution mass spectrometry. Compositional profiles of UVCB substances have been comprehensively evaluated through signature characteristics in the general mass spectrum, as well as relative abundance distributions at a molecular level (Castiblanco et al. 2020; Covas et al. 2020; Liu et al. 2020; Vanini et al. 2020). Comprehensive analysis of the chemical composition of UVCBs has provided information on the nature, genesis, and processes of a substance, therefore defining its intrinsic characteristics (Benassi et al. 2013; Mennito and Qian 2013; Oldenburg et al. 2014; Oldenburg et al. 2017; Orrego- Ruíz 2018; Silva et al. 2020; Walters et al. 2015; Wang et al. 2020). Additionally, the ability to study molecular trends has generated a novel understanding of the compositional transformations and variabilities within and across UVCB substances (Hosseini and Sachsenhofer 2021; Rocha et al. 2018; Silva et al. 2020). In summary, high-resolution mass spectrometry is a valuable tool in understanding the fate and behavior of UVCB substances (Jaggi et al. 2019; Li et al. 2022; Wozniak et al. 2019).

1.7. Application of ion mobility mass spectrometry for comprehensive assessment of UVCB substances

Ion mobility enables multidimensional measurements on the millisecond timescale through a post-ionization separation of ions as they travel across an inert gas under the influence of an electric field. The mobility measurements of each ion is directly correlated to their collisional cross section values, a descriptor of the ion's unique structural conformation (Dodds and Baker 2019). The fast screening capabilities and high sensitivity of ion mobility instruments coupled with the high-resolution of TOF-MS enables multidimensional information for a complete characterization of complex substances (Fernandez-Lima et al. 2009; Ibrahim et al. 2016). Combination of isomeric separation with high-resolution isobaric information is a promising aid to the deconvolution of UVCB substances (Fernandez-Lima et al. 2009; Grimm et al. 2017). Therefore, the application of ion mobility mass spectrometer (IMS-MS) is a valuable tool during the assessment and characterization of UVCB substances by defining "*individual chemical constituents, their structural features, and quantitative variation*" (ECHA 2020).

In addition to understanding the comprehensive characterization of UVCB substances, translation of the information to practitioners is another critical challenge. Petroleomic analysis has yielded molecular level information from high-resolution mass spectrometry data(Hsu et al. 2011; Marshall et al. 2010; Marshall and Rodgers 2004; 2008; Palacio Lozano et al. 2020; Santos et al. 2015). Chemometric approaches such as principal component analysis or hierarchical clustering have been applied to translate the molecular level information and to define the compositional trends and relationships between

samples (Hur et al. 2010). Principal component analysis aids the visualization of compositional variance by summarizing the dimensionality of a comprehensive profile (Corilo et al. 2013). Unsupervised classification of the compositional profiles through hierarchical clustering analysis allows one to visualize the commonality between substances in the generated clusters (Guillemant et al. 2019). Fowlkes-Mallow (FM) Index measures the similarity between clusters, to evaluate the concordance of experimental data-derived clustering to that of the pre-defined sample classification. FM index values can range from 0 (no correspondence) to 1 (perfect correspondence) (Fowlkes and Mallows 1983).

Novel analytical advances, such as IMS-MS presented here, have greatly increased our ability to access valuable information on the composition of UVCBs. The ability to characterize petroleum UVCB substances at a molecular level with multidimensional, high-resolution petroleomics addresses the major challenges encountered during risk assessment (Salvito et al. 2020). Therefore, through the implementation of petroleomics analysis within a comprehensive assessment framework to screen the chemical composition and complement other available information, IMS-MS can serve as a rapid and high-resolution tool to aid read-across and safeguard the practice of risk assessment. Through the application of non-targeted, high-resolution IMS-MS analysis we aim to shed light on relevant compositional information of UVCB substances previously thought unattainable. The work presented in this dissertation establishes a paradigm shift to the risk assessment of UVCB substances when applying read-across analysis providing information on "*structural similarity between substances which result in a likelihood that*

the substances have similar physicochemical, toxicological and ecotoxicological properties" (ECHA 2020).

1.8. Specific Aims

Throughout the work presented herein, we aimed to safeguard the hazard and risk assessment of UVCB substances with comprehensive chemical profiles. The overall objective is to develop high-resolution mass spectrometry methods for compositional characterization at a molecular level for comprehensive characterization, identification, and grouping of UVCB substances to safeguard proper hazard and risk assessment.

Specific Aim 1: To establish ion mobility spectrometry-mass spectrometry (IMS-MS) as a rapid and highly informative technique for the categorization of UVCB substances. Through this aim, the comprehensive compositional profiles acquired by high-resolution mass spectrometry were evaluated as qualitative information for geochemical categorization of crude oils. Crude oil samples from neighboring areas in the Gulf of Mexico region were analyzed using high-throughput IMS-MS and standardized GC-MS methods. The generated chemical profiles were used to group samples based on their geographical origin. The data was presented to practitioners through a table-top exercise to define the value and drawback of high-resolution IMS-MS as a decision-making tool.

Specific Aim 2: To define the composition of UVCB substances at a molecular level to define the degree of commonality across production and categories. In this aim, the molecular composition of petroleum refined products was deconvoluted through IMS-MS and petroleomic analysis. Accurate mass measurements, structural composition,

and relative abundances allowed to define the molecular composition and chemical profiles of each substance. The comprehensive characterization of the composition as a whole and its constituents then validated the intrinsic compositional variabilities and similarities between production cycles and across manufacturing categories.

Specific Aim 3: To characterize compositional transformations and their role on the fate and behavior of UVCB substances. With this aim, the role of photooxidation as a weathering process affecting the physical-chemical composition of the oil was investigated. By conducting an outdoor mesocosm experiment with a crude oil slick on seawater, we measured the direct effect of sunlight irradiation on the fate and behavior of the oil through IMS-MS analysis. Comprehensive molecular characterization of the water-soluble fraction allowed us to identify photooxidation products and describe physicochemical transformations as an effect of sunlight irradiation across time.

In summary, this dissertation provides a comprehensive understanding of the molecular composition of UVCB substances and defines the role of high-resolution mass spectrometry information in read-across approaches, therefore enabling risk assessment to better safeguard human health and the environment.

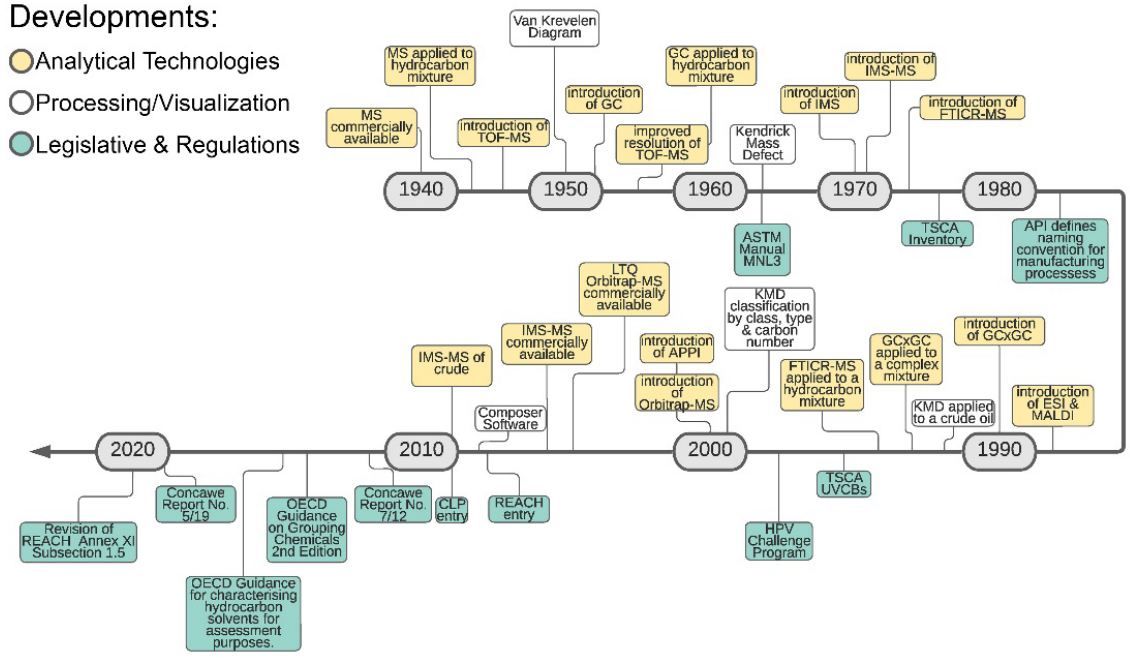


Figure 1.1 Evolution of the main developments in high-resolution analytical technology, petroleomic analysis and international regulatory frameworks regarding UVCB substances.

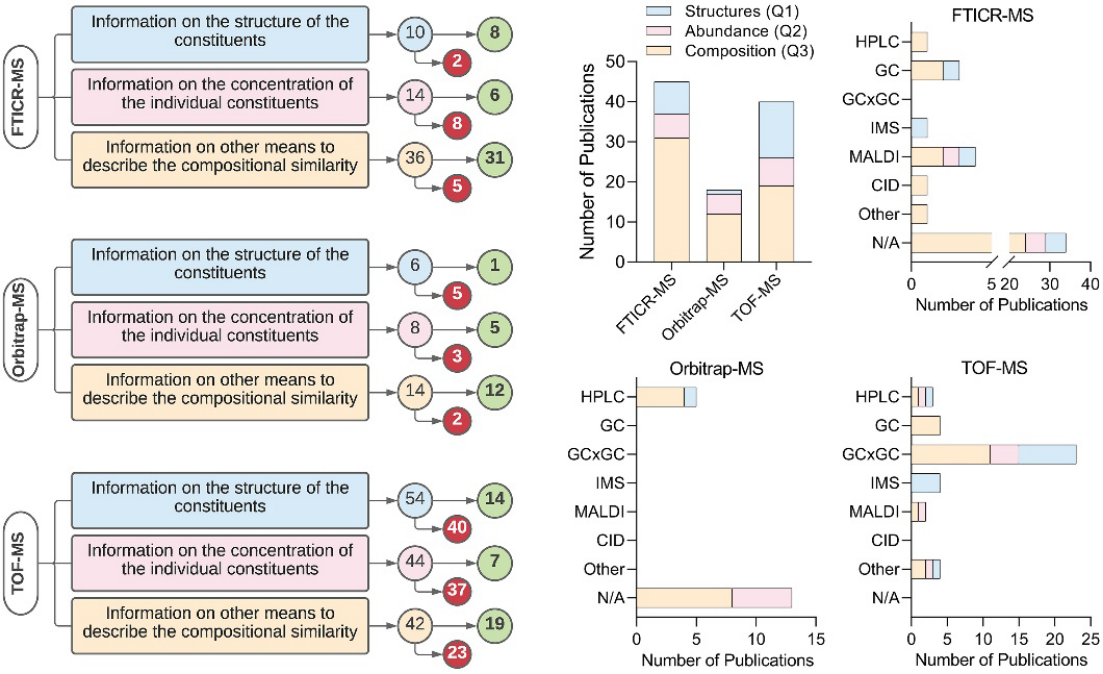


Figure 1.2 High-resolution mass spectrometry application throughout literature to address the regulatory needs established in the amendments of REACH Annex XI, S.1.5. (ECHA 2020).

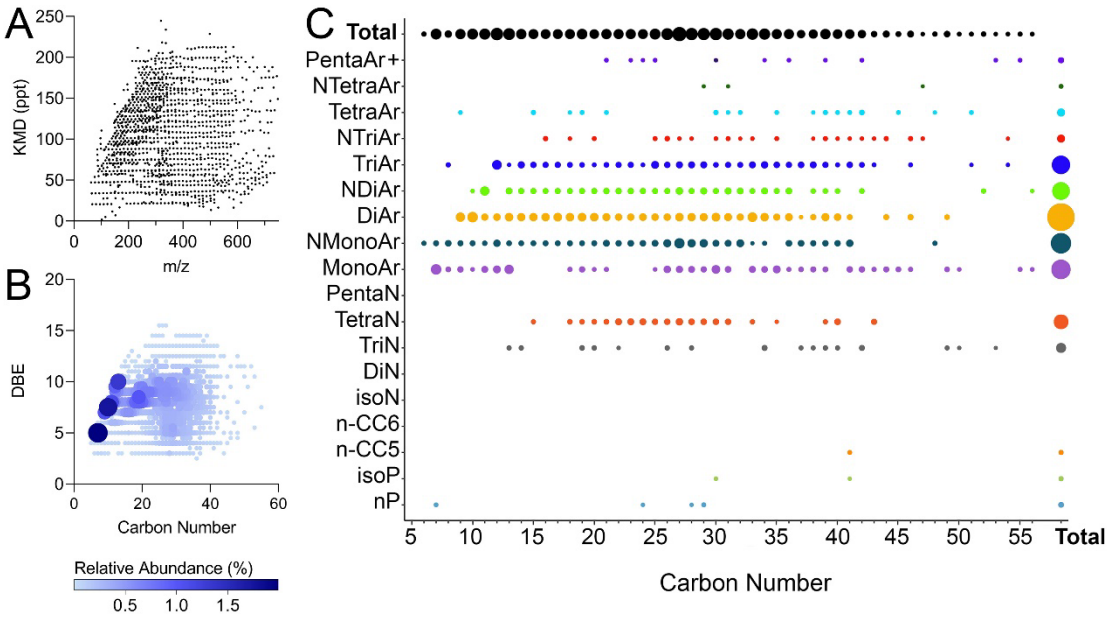


Figure 1.3 Petroleomic assessment of high-resolution mass spectrometry data. (A) Kendrick Mass Defect plot generating homologous series, (B) Double Bond Equivalent versus carbon number for hydrocarbon constituents and (C) hydrocarbon block assignment based on class and carbon range for a hydrotreated heavy paraffinic petroleum distillate.

2. A COMPARATIVE ANALYSIS OF ANALYTICAL TECHNIQUES FOR RAPID OIL SPILL IDENTIFICATION*

2.1. Overview

The complex chemical composition of crude oils presents many challenges for rapid chemical characterization in the case of a spill. A number of approaches are currently used to "fingerprint" petroleum-derived samples. Gas chromatography coupled with mass spectrometry is the most common, albeit not very rapid, technique; however, with limited resolution to resolve the complex substances in crude oils. This study examined the potential application of IMS-MS, coupled with chem-informatic analyses, as an alternative high-throughput method for the chemical characterization of crude oils. We analyzed 19 crude oil samples from on- and off-shore locations in the Gulf of Mexico region in the United States using both GC-MS (biomarkers, gasoline range hydrocarbons, and alkanes) and IMS-MS (untargeted analysis). Hierarchical clustering, principal component analysis, and nearest-neighbor-based classification were used to examine sample similarity and geographical groupings. We found that direct injection IMS-MS performed either equal or better than both GC-MS in the classification of the origins of crude oils. In addition, IMS-MS greatly increased the sample analysis throughput (minutes versus hours per

*Reprinted with permission from "A Comparative Analysis of Analytical Techniques for Rapid oil Spill Identification" by Alina Roman-Hubers, Thomas McDonald, Erin Baker, Weihsueh Chiu and Ivan Rusyn, 2020. Environmental Toxicology and Chemistry, 40, 1034-1049, Copyright 2020 by Wiley Online Library.

sample). Finally, a tabletop science-to-practice exercise, utilizing both the GC-MS and IMS-MS data, was conducted with emergency response experts from regulatory agencies and the oil industry. This activity showed that the stakeholders found the IMS-MS data to be highly informative for rapid chemical fingerprinting of complex substances in general, and specifically advantageous for accurate and confident source-grouping of crude oils. Collectively, this study shows the utility of IMS-MS as a technique for rapid fingerprinting of complex samples and demonstrates its advantages over traditional both GC-MS based analyses when used for decision-making in emergency situations.

2.2. Introduction

Large quantities of crude oil and petroleum refining products are released into the environment each year due to accidental spills or natural seeps; however, both exposure and hazard evaluation of these releases remain a formidable challenge to both the industry and the regulators (Laffon et al. 2016). One of the major confounding factors in oil spill response is the complexity of substance composition, making it excessively difficult to both characterize chemical constituents and identify their source (Bejarano and Michel 2016). Chemical analysis of petroleum has been an active area of the analytical chemistry for over 50 years and continuous advances in the methods for sample collection, extraction, analytical detection and data analysis have contributed to improvements in the accuracy of exposure characterization after oil spills (Aeppli et al. 2013; Corilo et al. 2013; McKenna et al. 2013; Onel et al. 2019; Stout and Wang 2007; Ventura et al. 2011; Wang et al. 2011; White et al. 2020). A number of novel analytical methods have been applied to these “petroleomic measurements” with the goal to achieve a more detailed chemical

characterization of the constituents, as well as derive characteristic signatures, or "fingerprints," of the particular oils or refining products (Fernandez-Lima et al. 2009; Marshall and Rodgers 2004; 2008). These methods typically couple several analyses together and the multi-dimensional datasets are then processed with various classification and regression methods to determine the similarities and differences in compositional patterns among samples (de Carvalho Rocha et al. 2017; Juahir et al. 2017; Onel et al. 2019).

Chemical fingerprinting of petroleum and other complex substances has traditionally relied on gas chromatography separations followed by high selectivity detection with mass spectrometry or flame ionization detection. For example, methods established by the United States Environmental Protection Agency 8051B "*Non-Halogenated Organics Using Gas Chromatography-Flame Ionization Detection"* is a common fingerprinting approach for qualitative forensic fingerprinting, and 8270 "*Semi-Volatile Organic compounds by Gas Chromatography-Mass Spectrometry"* is the employed for forensic fingerprinting of petroleum and complex mixtures. Both of these methods are widely used for characterization of complex samples and their individual constituents (Hantao et al. 2012). In addition, two-dimensional gas chromatography analyses have been shown to hold promise for oil fingerprinting (Aeppli et al. 2013; Ventura et al. 2011). However, while gas chromatography based analytical techniques provide useful and quantitative information, the chemical complexity of oil far exceeds the separation capacity of gas chromatography alone. In addition, gas chromatography methods are generally time-consuming (requiring at least an hour or more per sample) and

also face difficulties with identification of low abundance and/or high molecular weight compounds. Therefore, even though they are widely used for forensic fingerprinting of oil, gas chromatography methods are not considered to be most information-rich or fast techniques (de Carvalho Rocha et al. 2017; Fernandez-Lima et al. 2009; Marshall and Rodgers 2004; Onel et al. 2019; Stout and Wang 2007).

Several high-resolution mass spectrometry techniques have been also used for oil fingerprinting. Most used technique is Fourier-transform ion cyclotron resonance mass spectrometry (Corilo et al. 2013; McKenna et al. 2013; Purcell et al. 2007a; Purcell et al. 2007b). Other methods, such as X-ray fluorescence and attenuated total reflectance - Fourier transform infrared spectroscopy (White et al. 2020) and IMS-MS (Ibrahim et al. 2015; Santos et al. 2015) have also been explored for untargeted chemical fingerprinting of petroleum substances. IMS-MS is a rapid post-ionization technique that separates ions based on shape, charge and mass (Dodds and Baker 2019; Luo et al. 2020a). Furthermore, the ion mobility dimension allows the derivation of collisional cross section, a value for each chromatographic feature that is related to its rotationally-averaged area of interaction with the buffer gas in a drift tube (Dodds et al. 2020). IMS-MS may be coupled with different ionization sources, such as electrospray ionization (ESI), atmospheric pressure photoionization (APPI) or atmospheric pressure chemical ionization (APCI), depending on the molecular classes of interest. Previous studies with IMS-MS have shown promise in providing faster and information-rich content for complex substances; however, this technique is not as familiar to practitioners in oil spill response as gas chromatography-mass spectrometry (Fernandez-Lima et al. 2009; Grimm et al. 2017; Santos et al. 2015).

In this study, we tested the utility of IMS-MS as a high-throughput, untargeted screening method for characterization of crude oil's complex composition. While gas chromatography-mass spectrometry and other traditional techniques are used routinely for characterization the composition of key constituents in oils, we hypothesized that IMS-MS may be a sensible alternative for rapid analysis and identification of the origins of unknown samples in emergency situations. To test this, we performed a comparative analysis of IMS-MS and gas chromatography-mass spectrometry data for grouping of 19 crude oil samples from 6 distinct areas of the Gulf of Mexico region. The translational utility of each dataset was evaluated through a science-to-practice tabletop exercise with the stakeholders from the industry and regulatory agencies who were asked to make rapid assessment of the origins of two blinded samples.

2.3. Materials and Methods

2.3.1. Sample selection

Crude oil samples (n=19) from on- and off-shore locations in the Gulf of Mexico region were obtained from a repository of crude oil samples at Texas A&M University (Kennicutt II et al. 1992). Oil samples were stored at -20^{o}C until analyzed. The samples originate from two regions and three areas in the United States (Figure 2.1). Detailed geographical information about each sample is provided in Table S2.1. Sample labels correspond to the region and area of genesis. On-shore samples were identified by their State of origin (TX=Texas, LA=Louisiana, AL=Alabama) and a unique sample ID. Off-shore samples are identified with the letter "G" (for Gulf of Mexico), geographical area of

genesis (HI=High Island, ST=South Timbalier, EC=East Cameron) and a unique sample ID.

2.3.2. Sample preparation

To prepare the samples for analytical evaluation, the following solvents (all HPLC grade and ≥99.8% pure, Sigma Aldrich, St. Louis, MO) were used: toluene (CAS 108-88-3; Cat #34866), methanol (CAS 67-56-1; Cat #34860) and dichloromethane (CAS 75-09-2; Cat #34856). For GC-MS analyses, all oil samples were weighed and dissolved in dichloromethane (no precipitate was visible and it is assumed that the samples dissolved completely) to a final dilution of 1 mg of oil per 1 mL of dichloromethane. For the IMS-MS analyses, the same dichloromethane-diluted oil samples were used, but they were solvent exchanged from dichloromethane to 1:1 v/v toluene:methanol. (Grimm et al. 2017; Purcell et al. 2007a)

2.3.3. GC-MS instrumental parameters

The modified version of the established United States Environmental Protection Agency method 8270 was used for the analyses using an Agilent 7890 gas chromatograph (Agilent Technologies, Santa Clara, CA) interfaced with a Hewlett-Packard 5976 Mass Spectrometer (Agilent Technologies). HP-5ms Ultra Inert Column (30 m × 0.25 μm × 0.25 mm; Cat. #G3900-63001, Agilent Technologies) was used to chromatographically separate petroleum hydrocarbons. Instrumental operating conditions were as follows: mass range 45-500 m/z, split-less injector, injection volume of 2 μL, column flow 1mL/min, helium carrier gas. Initial temperature of the injection port was held at 250oC. The oven was initially set to 50°C with a hold time of 4 min, then the oven was

programmed at a rate of 6°C/min until it reached the final holding temperature of 300°C with a final hold time of 20 min. The petroleum hydrocarbons were determined using full scan mode (45 to 500 m/z). Full scan utilizes computer libraries to compare unknown analyte spectrums within the entire range of ions generated, providing information to resolve or confirm peaks qualitatively, pattern recognition, and structural elucidation (Wang et al. 2006).

2.3.4. GC-MS data analysis

Full-scan chromatograms were analyzed using ChemStation (Agilent) software for identification and detection of individual n-alkanes, selected isoprenoids and parent and alkylated polycyclic aromatic hydrocarbons (PAH). See the list of compounds analyzed by gas chromatography-mass spectrometry and their abbreviations in Table S2.2. Raw data was exported to a table containing the specific abundances of each sample. Each sample's data was normalized as follows - each peak's area was divided by the sum of the individual peak areas in that sample and multiplied by 100.

2.3.5. IMS-MS instrumental analysis

Samples were directly injected into the Ion Mobility Spectrometry Quadrupole-TOF Mass Spectrometer (Agilent, model G6560A; serial# SG1711C002) at 50 µL/min. Each sample was analyzed in triplicate through independent injections of the same sample, due to the instrument's high-throughput capacity for running and analyzing samples. Initial ionization was carried out with an APPI source (Agilent, model# G1917C). Instrumental and source parameters were as follows: APPI positive ion mode, sample analysis time 1.5 minutes; source parameters: gas temperature 325oC, vaporizer 350oC,

drying gas 10 l/minute, nebulizer 30 psi, VCap 3000, fragment 400V, 110 RF Vpp 750. The following acquisition parameters were defined in each instrumental run: mass range 50-1,700 m/z, frame rate 1 frames/s, IM transient rate 18 transients/frame, max drift time 60 ms, TOF transient rate 600 transients/IM transients, trap fill time 20,000 µs and trap release time 300 µs. QTOF parameters were as follows: firmware version 18.723, Rough Vac 2.71Torr, Quad Vac 3.68×10-5 Torr, TOF Vac 3.47×10-7 Torr, drift tube pressure 3.940 Torr, trap funnel pressure 3.790 Torr, chamber voltage 5.96 µA and capillary voltage 0.076 µA. Data was obtained using the MassHunter Acquisition software (Agilent, v.08.00).

2.3.6. IMS-MS data analysis

The IMS-MS raw files were uploaded to the MassProfiler software (Agilent, v.08.00) for feature analysis and identification. Raw data were recalibrated using MassHunter Browser Acquisition Data software (Agilent, v.08.00) to derive nitrogen gas-filled drift tube collisional cross section ($^{DT}CCS_{N2}$) values of the detected features. Data filters were set at abundance ≥1,000 and Q-Score >75. The individual and grouped feature data matrices containing *m/z*, drift time, $^{DT}CCS_{N2}$, and abundance for each sample were then exported to Microsoft Excel for further evaluation. All IMS-MS features were first combined across the samples, comprising a dataset of 23,639 features (referred to as "IMS-MS all features"). Then, these data were further filtered based on observed frequency (>1) among all samples and among triplicates (2/3) within the same sample. This step yielded a total of 4,133 features across the samples (referred to as "IMS-MS untargeted" dataset). The final filtering step was based on selection of the

features that had predictions of their molecular formula by Agilent MassHunter MassProfiler software; this dataset included 939 features across all samples (referred to as “IMS-MS targeted” dataset). Abundances were then normalized to log10(abundance+1) fraction of the features in each sample.

2.3.7. Petroleum biomarker assay

The biomarker data used in this study was originally reported in (Kennicutt II et al. 1992). Approximately 100 mg of crude oil was spiked with a surrogate mixture (5β-cholane, d10-phenanthrene, ad d12-chrysene), and the mixture was then fractionated into its saturated and aromatic fractions using high performance liquid chromatography, as detailed elsewhere (McDonald and Kennicutt II 1992). The saturated fraction was then used for a gas chromatography-based petroleum biomarker analyses using the Hewlett-Packard 5890 gas chromatography coupled with Hewlett-Packard 5790A MSD. The compounds that were used for generating the chemical fingerprint (biomarker ratios) data in samples from the gas chromatography- mass spectrometry data were diasteranes, steranes (C27-C29), hopanes, tricyclics and moretanes (see Table S2.3-8 for the data).

2.3.8. Data integration and hierarchical clustering

Data matrices from the gas chromatography- mass spectrometry (n-alkanes and biomarkers) and IMS-MS analyses were then used to carry out data integration and hierarchical clustering to group crude oil samples based on their chemical profile. For each dataset, several different data transformations were evaluated, including log-transformation, normalization to percent abundance, and log-transformation after normalization. Unbiased hierarchical clustering analysis was performed using *hclust*

package in R studio (version R-3.6.2) with the default "complete linkage" method for similarity. The product of this analysis was a dendrogram summarizing the correlation among the chemical profiles for the samples analyzed with both analytical methods. Principal component analyses were then utilized to translate data matrixes and determine the variance between the chemical profiles among samples in order to evaluate their genesis. These analyses were carried out using *prcomp* and *scatterplot3d* packages in R. For nearest neighbor analyses, a distance matrix was calculated using the *stats* package in R applying either Euclidean distance (all methods except "IMS-MS all features" dataset) or a binary metric (for the "IM-MS all features" dataset). Then, for each sample, the nearest neighbor was defined as the sample with the smallest distance, and an accurate classification was defined as the sample having the same geographic region as its nearest neighbor. The percentage of accurate classifications (i.e., correctly identified with respect to the region and area of genesis) was calculated for each dataset and data transformation combination. In addition, the number of incorrect classifications across all dataset/ transformation combinations was tabulated.

To evaluate the outcome of the clustering, we derived a quantitative metric to assess the correspondence of the outcome to the original groups of each sample. The details of the unsupervised analysis workflow are described elsewhere (Onel et al. 2019). The Fowlkes-Mallows (FM) index (Fowlkes and Mallows 1983), a measure of similarity of two clusters, was calculated to enable quantitative comparative assessment between groupings achieved using each dataset to the known origins of each sample. The higher the FM index, the more similar the grouping to the *a priori* determined grouping as shown

in Table S2.1. The FM index ranges from 0.0 (no correspondence) to 1.0 (perfect correspondence). One-sided p-values for the FM index (using the null hypothesis of random assignment) were obtained using a standard z-statistic (Fowlkes and Mallows 1983) that compares the observed value to the null expectation.

2.3.9. Table-top exercise

A half-day meeting was held with a diverse group of oil spill response experts from petrochemical companies (Chevron Phillips Chemical Company, ConocoPhillips, ExxonMobil, Occidental Petroleum, Shell, Chevron), and state regulatory agencies (Texas General Land Office and Texas Commission on Environmental Quality). Pre-table top exercise orientation of the participants consisted of a series of short presentations covering the overall goals for the table top exercise, an overview of the traditional analytical methods and IMS-MS, and a summary of the data to be evaluated. The participants were asked to consider a scenario in which several different forensic fingerprinting analyses were used to identify the source of the recently spilled oil by comparing two blinded crude oil samples to a library of crude oils from known locations. Specifically, the data package presented to the participants included information on the samples from six on- and off-shore Gulf of Mexico oil production regions of which 2 samples were blinded with respect to their origins. The identities of the two blinded samples were US_AL_L7256 (sample 1) and US_G_HI_L7291 (sample 2). Both “traditional” data on bulk composition (such as selected sterane, triterpene and hopane peaks, selected gasoline range hydrocarbons, data on n-alkane and isoprenoid alkanes, full scan GC-MS chromatograms, and tabular data

representative of n-alkane abundance), and "new" data (IMS-MS 2-dimensional plots and clustering diagrams) were provided.

The participants were divided into three groups of 5 individuals, all representing different companies/agencies, and were given 60 minutes to review, discuss and compare the data and predict the genesis of the unknown samples. Each group reached an independent conclusion on the origins of two blinded samples and also provided answers to the following general questions: (i) How easy/difficult was it to identify the "unknown" samples using petroleum biomarker, GC-MS, or IMS-MS data? (ii) How did the different data types compare in terms of their performance/quality/timeliness? (iii) What is needed to increase the use and application of IMS-MS data to facilitate the transition from research to practice? and (iv) Overall, what is your impression of the IMS-MS and its proposed application to oil spill response?

2.4. Results

The oil samples selected for this study (Supplemental Table 1) represented distinct oil-producing geographical regions in the Gulf of Mexico from both on- and off-shore locations (Figure 2.1). For each region, samples from wells located in close proximity to each other were included to test how well GC-MS and IMS-MS perform for rapid characterization of the region/area of origin.

2.4.1. Compositional analysis

Traditional methods of petroleum fingerprinting rely on petroleum biomarkers, organic molecules that are highly conserved and representative of the geographical origins and genesis of each oil sample (Radovic et al. 2012). The presence of petroleum

biomarkers, such as steranes (*m/z* 217) and triterpanes (*m/z* 191), indicates historical geological organic matter and deposits in crude oil; these molecules are most useful geological information for forensic fingerprinting (El-Sabagh et al. 2018). Oil sample biomarker data were available for 15 of 19 crude oils samples included in this study; for samples from TX and LA, regular steranes and triterpanes were below detection limit (Supplemental Tables 3-6). The relative fractions of regular steranes and triterpanes, shown in Figure 2.2A, indicate strong qualitative and quantitative similarities among the compositional of the samples in each region. For example, the characteristic feature of the on-shore AL samples was the relatively low abundance of diasteranes and high abundance of C29 steranes, a biological marker indicating carbonate rock source. The off-shore samples, as expected for the marine organic source oil, contained a relatively high abundance of diasteranes and high abundance of tricyclic terpanes. Among triterpanes, hopanes were the most predominant species in all samples.

The data on sample-specific patterns in C27, C28, and C29 regular steranes (Figure 2.2B) was used to distinguish the sources of organic matter in each sample. The ternary diagram gives some indication of the source-specific differences among samples. On-shore AL samples exhibited low abundance of C27 and C28 steranes, and a considerably higher abundance of C29 steranes, features that provided some separation from the off-shore samples. The off-shore oil samples were characterized by abundance of C27 steranes and lower abundance of C28 and C29 steranes, indicating a marine organic source; however, it was largely impossible to distinguish the individual regions among the off-shore samples with these data alone.

GC-MS full scan chromatograms (Figure 2.3) showed the characteristic patterns of the chemical constituents in the samples. The GC-MC analysis used in this study was carried out to simulate a typical rapid analysis approach for identification of the crude oil's fingerprint. These chromatograms demonstrated some location-specific differences in the n-alkane distributions and the extent of the unresolved complex mixture (UCM) "hump". Semi-quantitative datasets were derived from GC-MS runs for gasoline-range hydrocarbons (Table S2.9 and Figure S2.1-3), n-alkanes (Table S2.10 and Figure S2.4-6), the GC-MS alkanes data (Table S2.11 and Figure S2.7-9), or parent and alkylated PAHs (Table S2.13 and Figure S2.10-12).

Visual examination of the chromatograms and data plots is a common method for evaluation of the hydrocarbon patterns. In the gasoline-range hydrocarbons dataset (Table S2.9 and Figure S2.1-3), some characteristic patterns were evident among samples from different regions and areas. For example, benzene was highly abundant in on-shore samples from TX; 2-methylhexane and 3-methylhexane were highly abundant in samples from LA; and n-C3 to n-C6 range hydrocarbons were highly abundant in samples from AL. The off-shore samples were largely similar to each other and no characteristic patterns were observed to distinguish among the three areas, which is largely expected because the oil originates form the same source rock in the Gulf of Mexico region.

The data on composition of C3-C32 alkanes (Supplemental Table 10 and Figure S2.4-6) is useful to establish whether sample degradation may have occurred, such as the relatively low abundance of n-alkanes compared to isoalkanes (Wang and Stout 2010). Specifically, we found that some of the off-shore samples from HI area (L5356, L5361,

L7281, and L7291) had a relatively low n-alkane to isoalkane ratio, a sign of possible degradation in these samples which is concordant with a greater UCM hump in these samples as compared to the other two samples from this area (L5346, L5351). Uneven sample degradation in the HI set may present challenges with correct grouping of these samples based on these data alone. However, these data are useful for distinguishing between on- and off-shore samples. Specifically, the pristane-to-phytane ratio is characteristic of the crude oil's source rock. On-shore samples exhibited low pristane/phytane ratio and a preponderance of C5-C12 n-alkanes, while the off-shore samples displayed higher pristane/phytane ratio.

The graphs generated from the GC-MS alkanes data (Table S2.11 and Figure S2.7-9) demonstrate that some of the area-specific samples did have similar analytical profiles. For example, the hydrocarbon profiles of on-shore samples from AL and LA were very similar within their respective group, but different between groups. Similarly, ST area off-chore samples were very similar to each other. However, these data were largely uninformative with respect to grouping of the samples from other areas. While some of the HI area samples may have been partially degraded which would be explaining the dissimilarities in this dataset, other area samples (EC and TX) showed no evidence of degradation. Overall, the data from GC-MS were informative for oil fingerprinting in our dataset, but it is also evident that for full confidence in grouping, further analytical characterization of target analytes for a detailed compositional analysis may be required.

Evaluation of the GC-MC chromatograms for peak areas of the selected PAHs (Table S2.12 and Figure S2.10-12) showed presence of characteristic petrogenic 2- and 3-

ring PAHs in on-shore samples from LA and AL, the fingerprints generated from TX samples demonstrated high abundance of 2-ring PAHs. Off-shore samples exhibited predominance of 2-ring PAHs and traces of 3-ring PAHs.

Next, we examined the utility of IMS-MS data for sample grouping (Figure 2.4). Representative 2-dimensional plots (selected from 3 technical replicates obtained for every sample because of the speed of the analysis) show drift time versus mass-to-charge (*m/z*) nested spectra derived from APPI-assisted sample ionization analyses of the samples (see Table S2.13-14 for the numerical data). APPI was utilized here to focus on the non-polar compounds (e.g., polycyclic aromatic compounds, PAC). Visual inspection of the IMS-MS nested spectra most clearly demonstrated the differences between samples from on- and off-shore regions. The compositional complexity of the on-shore samples was greater than that of off-shore samples, as evident from the greater abundance of the high molecular weight features in the former. In addition, the 2-dimensional nested spectra also illustrated visual trendline differences; the on-shore samples have 3 clear trendlines, but the off-shore samples only had 1 main trendline. A difference in the feature abundances was also noted for the different trendlines based on the region, allowing visual separation and easy grouping.

2.4.2. Chemical-data integrative compositional fingerprinting

While the visual inspection of the raw spectra/chromatograms and plots from some of the chemical components constitutes a useful approach to data analysis and sample characterization, a formal statistical evaluation was conducted to determine the ability of each dataset to distinguish among regions and areas. First, we used the data from both GC-

MS and IMS-MS to conduct hierarchical clustering analysis (Figure 2.5) which is an unsupervised method for grouping samples. With GC-MS alkanes data, correct grouping was evident for on-shore samples from AL and LA, and off-shore samples from ST. However, these data did not distinguish between on- and off-shore samples; in addition, the samples from other regions were interspersed. With GC-MS PAH data, in many regions, samples clustered together; however, on- and off-shore samples were interspersed and LA samples did not cluster together. Using the biomarker data, this analysis achieved the separation between on- and off-shore samples. Among on-shore samples, AL samples clustered together, but LA and TX samples were interspersed. Among off-shore samples, ST samples formed a cluster, but other samples did not group correctly. When IMS-MS data were used, regardless of whether the untargeted or targeted lists (Supplemental Table 14) were used, all samples were grouped correctly into their region and then area.

A quantitative metric for how well unsupervised clustering corresponded to the actual groupings (Table S2.1), a Fowlkes-Mallows (FM) index was calculated (Table S2.15). For the ability of these data to group samples into two regions (off- or on-shore), the highest FM index of 1.0 resulted from the clustering using IMS-MS (either untargeted or targeted) data. The biomarker data yielded FM index of 0.64 ($p=0.04$), while GC-MS data were uninformative (for alkanes: FM=0.48, $p=0.37$; for PAH: FM=0.51, $p=0.42$). When grouping into 6 specific areas was examined, again, IMS-MS data were most informative (FM=0.56-0.66, $p<0.001$), followed by full scan GC-MS data (FM=0.34, $p=0.01$). The biomarkers data classification was not significantly different from random assignments of samples into groups (FM=0.28, $p=0.06$).

Next, we used principal component analysis as an alternative widely-used unsupervised data visualization method (Figure 2.6). Similar to the outcome of the hierarchical clustering analysis, data from GC-MS, either alkanes or PAH, demonstrated inadequate separation among areas and regions. When using all features IMS-MS dataset, we found distinct grouping of the samples between on- and off-shore locations; however, even in the first three principal components the principal component analysis showed considerable overlap among the samples from the off-shore areas. The targeted IMS-MS dataset was superior in distinguishing between both areas and regions.

Lastly, we used a nearest-neighbor classification analysis (Table 2.1) to evaluate the ability of the different datasets to predict the region and area from which each sample originated. With GC-MS alkanes data (Table S2.11, Figure S2.13-16), on-shore samples from AL and off-shore samples from ST and EC can be classified into their respective region/area with high accuracy; these data were far less informative for classification of the samples from other areas. Interestingly, GC-MS PAH data (Table S2.12, Figure S2.17-20) afforded more accuracy in classification, in the range of 74 to 95% with only one mis-classified sample (LA L7186) for the log-transformed abundance data. The biomarkers data (Table S2.-8, Figure S2.21-22) had comparable accuracy to that of GC-MS alkanes data; off-shore HI samples L5346 and L7291 were most challenging to classify accurately. Gasoline range hydrocarbon and % composition alkanes C3-C32 datasets (Table S2.9 and 10, Figure S2.23-24 and 25-26, respectively) were least accurate in classification accuracy, mis-predicting about half of all samples. Interestingly, when the data acquired through GC-MS (Table S2.3-11, Figure S2.27-28) were used, higher accuracy of

classification (84%) was achieved, but off-shore HI samples L5346, L5356 and L7291 remained difficult to classify. IMS-MS data were far more informative for accurate classification of these samples. Most of the datasets or data transformation methods yielded perfect classification accuracy (Table S2.13-14, Figure S2.29-37).

2.4.3. Science-to-Practice translation

To determine whether the data and analyses detailed above can be used by practitioners responding to an oil spill, we conducted a table top exercise with the simulated case scenario of the oil pipeline burst. Experts from oil industry and government agencies with expertise in oil spill response were invited to participate in this exercise. The participants were given a brief tutorial on the analytical methods used to generate the data and then asked to review the totality of the information and identify the geographical origins of two blinded crude oil samples (sample 1: US_AL_L7256 and sample 2: US_G_HI_L7291) by comparing their analytical profiles to a library of profiles of crude oils from known locations (the remaining 17 samples, Table 2.2). Participants were divided into three groups balanced with respect to their employer (industry vs government and ensuring that no individuals from the same company/agency were in the same group) to ensure the diversity of experiences, opinions and perspectives. Each group had to review all data, identify the samples based on each dataset separately, and come up with the recommendations in 60 minutes. Scientists with the knowledge of the analytical techniques were available to all groups to answer technical questions that arose during the discussion and decision-making.

All three groups correctly identified sample 1 as a sample from an on-shore AL area. Interestingly, both GC-MS and IMS-MS data were found to be equally informative for this sample's identification, because correct identifications were made using each dataset independently. However, sample 2 presented a more formidable challenge. All groups correctly predicted that this sample originated from the off-shore region; this outcome was achieved with either dataset. Traditional GC-MS data, either petroleum biomarkers or hydrocarbons, were found to be least informative by the participants; no group was able to correctly identify the sample's origins with confidence, except for group 1 with hydrocarbons GC-MS data. When using IMS-MS data, all three groups correctly identified this sample's origins as from the off-shore HI area.

2.5. Discussion

Crude oils are fossil fuels generated under immense pressure and high temperatures from deposited layers of the remains of prehistoric organisms. Because of the variability in the environment and type of deposits, crude oils have both complex and variable composition that depends on their geographical origins. Crude oil samples have a unique chemical fingerprint due to the geological conditions and the age of formation (Pang et al. 2015; Wang et al. 2006). The inherent chemical complexity and variability of crude oils create a unique challenge for their source characterization. Traditional analytical methods, primarily GC-MS, have been used widely for chemical fingerprinting of petroleum-derived samples by measuring n-alkanes, acyclic isoprenoids, PAC (including alkylated and sulfur-containing substances), as well as biomarkers (Daling et al. 2002; Stout and Wang 2007). High resolution mass spectrometry, such as IMS-MS-based

method, has been used to improve both analytical efficiency and resolution through incorporation of feature separation by shape and size (Grimm et al. 2017; Ibrahim et al. 2015; Ponthus and Riches 2013).

In this study, we aimed to compare and contrast traditional GC-MS-based petroleum fingerprinting to that performed using IMS-MS. Specifically, we evaluated the ability of each technique to identify the differences among, and similarities within, oil samples from several geographically close regions/areas of production. Petroleum biomarkers are commonly used for distinguishing the origins of oil samples because these compounds originate from once-living organisms and they persist in oil spills and in refined products (Peters et al. 2005). Biomarker-based fingerprinting of oil is mostly successful when applied to samples from distant locations (Hansen et al. 2007; Peters et al. 2005). In our case, oil samples were from the locations that were relatively geographically close, even if they were still distinct in terms of their formation and age (Table S2.1). Thus, even though we addressed a more difficult question of distinguishing among the samples from neighboring regions and areas, this challenge is not unusual in real-life oil spill response. Indeed, we observed some chemical similarities in biomarkers' distribution and abundance in crude oils originating from on- or off-shore regions (Kennicutt II et al. 1992). Biomarker content revealed most pronounced differences in sample composition between two regions (on- and off-shore); however, these data were not informative for distinguishing among the individual areas within each region. Similarly, the chromatograms generated by GC-MS did reveal the samples' chemical complexity. Still, the data on n-alkanes, isoprenoids and the UCM hump were not

sufficient to provide sufficiently detailed characterization of crude oil's composition, limiting the ability to distinguish the samples and identify their source. While we show the challenges in using conventional analytical methods for rapid assessments, further characterizations with time-consuming follow-up analyses may provide additional data that can aid in source identification.

In contrast to the traditional data, IMS-MS dataset demonstrated its utility for classification and grouping of all samples with higher degree of confidence. Prior studies (Grimm et al. 2017; Santos et al. 2015) of oil or products of petroleum refining utilized ESI-coupled IMS-MS, which is best suited for identification of heteroatoms (Purcell et al. 2007b; Zhan and Fenn 2000). In this study, we used APPI(+) IMS-MS to focus on the hydrocarbon molecules of higher carbon unsaturation than those produced by ESI (D'Andrilli et al. 2010). Indeed, IMS-MS chemical profiles were more effective in breaking down the complexity of crude oils for qualitative characterization and grouping. In addition, the ability to run technical replicates because of the short run times (less than 1 minute/sample), provided additional confidence in detection of the individual features in each oil samples. Not only two-dimensional IMS-MS spectra showed characteristic signature patterns for specific regions, with consistent distribution and abundance of constituents, but also the inclusion of feature separation by shape, size and mass allowed for detection of hydrocarbons that may have been previously unresolved due to co-elution in GC. Indeed, the data from IMS-MS were highly informative in the nearest-neighbor classification analysis where all datasets, either full feature list, or a reduced dataset with only features with predicted molecular formula, yielded perfect classification accuracy.

This finding is especially noteworthy because other datasets had challenges with classifying several samples correctly and it also appeared that there may have been some challenges with sample integrity.

The advantages of IMS-MS for rapid oil identification notwithstanding, we note that this study did not attempt detailed feature identification. Even though for over 900 chromatographic features in IMS-MS dataset a molecular formula could be assigned, verification of those assignments and precise feature identification still remain to be major technical challenges for the use of this and other untargeted analytical techniques for characterization of complex petroleum substances. Therefore, it is difficult to determine which individual features were most informative for classification. In addition, this study used a limited number of samples that may have handicapped our ability to classify them; however, this study's number of samples is very similar to that used by other groups to fingerprinting and source identification of oil (Corilo et al. 2013; de Carvalho Rocha et al. 2017; Ventura et al. 2011).

Another challenge with incorporating new analytical methods into practice of oil spill response is the lack of familiarity with these techniques in the industry and government agencies (National Research Council 1999; 2014). The GC- flame ionization detection and GC-MS methods are used widely and are familiar to all stakeholders, there is a greater operational experience with using these during spills over past few decades (Daling et al. 2002). Confidence in their performance and utility has been developed not only through various government guidelines, but also through collaborative studies that aim at both method standardization and building confidence among the stakeholders

(Faksness et al. 2002). The primary goal of such collaborations is to demonstrate the potential for a particular methodology to be both technically feasible and to generate scientifically defensible data useful in oil spill identification. The ability of a particular technique to distinguish qualitatively similar oils from a spill and any available candidate source is critically important for future implementation of the technology into practice.

Indeed, this study aimed to achieve science-to-practice translation of the benefits afforded by IMS-MS as a potential analytical method for oil spill response. We conducted a stakeholder workshop that included both traditional and IMS-MS data on the same samples as the first step toward qualification of this technique for oil spill response. We invited a diverse group of relevant practitioners, the participants represented different companies and government agencies and had varying levels of technical expertise and may have had different perspectives on oil spill response. Interestingly, not only all of the participants recognized the challenges with current "best practice" in forensic fingerprinting of complex substances such as crude oils, but also, some biased views notwithstanding, all of them were open to learning about new technologies.

In a relatively short time, about 60 minutes, the participants were able to evaluate different datasets. Both graphical and numerical data were provided; because of the compressed timeline, most of the discussion revolved around graphical data and clustering analysis. Two blinded samples were selected to represent a spectrum of difficulty with respect to identification. The on-shore AL sample was correctly identified by all groups with each dataset. The exercise of reviewing the data for the "easy" sample allowed participants to become comfortable with the data and also to compare and contrast the

information among datasets. The “difficult” off-shore HI sample presented a formidable challenge for identification with traditional data, but not IMS-MS data. Interestingly, the participants were consistent in their classification based on IMS-MS across all three sub-groups.

In addition to the exercise of sample identification, the participants were asked to comment on several general questions (Table 2.3). The responses provide additional insights into the strengths and weaknesses of each methodology, as well as they highlight the needs for additional IMS-MS method development before application in oil spill response. First, the participants found IMS-MS data most informative overall, especially with respect to the ability to appreciate patterns in features on the 2D plots and sample groupings on the clustering diagrams. Lack of clear groupings among named samples based on the data from traditional methods was a considerable impediment to reaching confident conclusions about the identity of blinded samples. Still, the participants commented that the information from each technique separately was not as powerful as the combination of the data form different sources. Second, the participants identified the speed of IMS-MS analysis, as compared to other data presented to them, to be a clear advantage. The short sample run times allowed for the analysis of technical replicates; the evidence that the replicates cluster closely with each other was a considerable positive factor in assigning greater confidence to IMS-MS data as compared to other datasets that lacked replicates. Still, the participants were unsure as to how the data quality is to be judged for the IMS-MS data as this technique is least known and the criteria assessing the quality of the features was found somewhat difficult to understand through a short

presentation. The fact that GC-MS data are more regulatory acceptable and familiar to the practitioners was mentioned as an advantage. As a follow up to this sentiment, the third question pertained to defining the barriers for the use of IMS-MS, or any other novel analytical technique, for oil spill response. The participants were uniform in their advice to prioritize feature identification, through the use of standards and other chemometric analyses such as Kendrick Mass Defect (Hughey et al. 2001), and investigation of the weathered oil samples. The analysis of reproducibility and lab-to-lab comparisons were also deemed important before this technology may become more widely accessible through contract laboratories. Lastly, the participants were very positively impressed with IMS-MS as a technique that is suitable for oil spill response. Again, the rapid analysis time was the primary identified benefit that would allow for this technique to be used as a method for triage of samples and focusing subsequent analyses with more traditional methods for confident feature identification. The combination of IMS-MS with other analytical techniques was identified as the most likely path to early adoption.

Overall, this study demonstrates that IMS-MS methodology is useful for grouping complex crude oils based on their genesis, even though it is currently insufficient for comprehensive feature identification and quantitative analysis. Solutions to address these challenges are the analysis of a wider range of chemical standards, a process that will enrich current IMS-MS libraries (Metz et al. 2017) with $^{DT}CCS_{N2}$ information for feature identification. The use of additional ionization sources, such as ESI and APCI, is needed for deeper characterization of the heteroatoms and aliphatic hydrocarbons in oils. Finally,

the application of IMS-MS to source identification of weathered oil samples should be the focus of future studies.

2.6. Acknowledgements

This work was supported, in part, by grants from the National Institutes of Health (P42 ES027704) and the National Academies Gulf Research Program (#2000008942). A.T. Roman-Hubers was supported, in part by a training grant from the National Institutes of Health (T32 ES0226568). The authors wish to thank the following individuals for participation in the stakeholder workshop: Sophie Jia (Chevron Phillips Chemical Company), Danile Arrieta (Chevron Phillips Chemical Company), Bradley Clark (ConocoPhillips), Justin Mellen (ConocoPhillips), Timothy Nedwed (ExxonMobil), Thomas Parkerton (ExxonMobil), Garret Choquette (Occidental Petroleum Corp), Hua Shen (Shell), Jeffrey Charlap (Chevron), Brent Koza (Texas General Land Office), Sabine Lange (TCEQ), Michael Honeycutt (TCEQ) and Kelly Scribner Tuttle (CTEH).

Table 2.1 Nearest-neighbor classification analysis to evaluate the ability of the different datasets to predict the region and area from which each sample originated. See Table S2.1 for detailed information on each sample.

Dataset	# of Features	Data Transformation	Distance Metric[e]	Percent Correct	US_AL_L7251	US_AL_L7256*	US_AL_L7266	US_LA_L7146	US_LA_L7186	US_TX_L8381	US_TX_L8386	US_G_EC_L6576	US_G_EC_L6581	US_G_HI_L5346	US_G_HI_L5351	US_G_HI_L5356	US_G_HI_L5361	US_G_HI_L7281	US_G_HI_L7291**	US_G_ST_L5001	US_G_ST_L5026	US_G_ST_L5031	US_G_ST_L5036
GC-MS (Abundance)	37	None	E	63	T	T	T	T	T	F	F	F	F	F	T	F	T	F	T	T	T	T	T
GC-MS (Abundance)	37	Log	E	74	T	T	T	T	T	F	F	T	T	T	T	F	F	F	T	T	T	T	T
GC-MS (Frac)	37	Norm.	E	74	T	T	T	T	T	F	F	F	T	T	T	T	T	F	F	T	T	T	T
GC-MS (Frac)	37	Norm.& Log	E	68	T	T	T	T	T	F	F	T	T	F	F	T	T	F	F	T	T	T	T
Biomarkers (All)[a]	78	None	E	63	T	T	T	T	F	F	F	T	T	F	F	F	T	T	F	T	T	T	T
Biomarkers (All)[a]	78	Log	E	89	T	T	T	T	T	T	T	T	T	F	T	T	T	T	F	T	T	T	T
Gasoline Range Hydrocarbons[a]	27	None	E	58	F	T	T	T	T	T	T	F	F	F	F	T	T	F	F	T	T	T	F
Gasoline Range Hydrocarbons[a]	27	Log	E	47	F	T	T	T	T	T	T	F	T	F	F	F	T	F	F	T	F	F	F
%Composition Alkanes C3-C32[a]	37	None	E	42	F	T	T	T	T	F	F	F	T	F	F	F	F	F	F	T	T	T	F
%Composition Alkanes C3-C32[a]	37	Log	E	63	T	T	T	T	T	F	F	T	T	T	T	F	T	T	F	F	F	T	F
GC-MS (Biomarkers)	115	Norm.	E	68	T	T	T	T	F	T	F	T	T	F	F	F	T	T	F	T	T	T	T
GCMS (Biomarkers)	115	Norm.& Log	E	84	T	T	T	T	T	T	T	T	T	F	T	F	T	T	F	T	T	T	T
IMS-MS (All)[b]	23,639	None	B	100	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
IMS-MS (All)[b]	23,639	Log	E	89	T	T	T	T	T	F	F	T	T	T	T	T	T	T	T	T	T	T	T
IMS-MS (All)[b]	23,639	Norm.& Log	E	89	T	T	T	T	T	F	F	T	T	T	T	T	T	T	T	T	T	T	T
IMS-MS (Targeted)[c]	939	None	B	100	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
IMS-MS (Targeted)[c]	939	Log	E	100	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
IMS-MS (Targeted)[c]	939	Norm.& Log	E	100	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
IMS-MS (Untargeted)[c]	4,133	None	B	100	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
IMS-MS (Untargeted)[c]	4,133	Log	E	100	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
IMS-MS (Untargeted)[c]	4,133	Norm.& Log	E	100	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
# of False Classified:					3	0	0	0	2	9	10	5	2	9	6	8	2	7	10	1	2	1	4

T = Nearest neighbor is from the same geographic region.

F = Nearest neighbor is from a different geographic region.

a, Datasets were already normalized as percentages.

b, Each technical replicate was treated as a separate sample.

c, Data were filtered and with technical replicates averaged.

d, Data transformation type: Norm., normalized; Log, log 10 transformed.

e, Distance metric: E, Euclidian; B, binary.

*, Sample #1 for which its identity information was withheld from the tabletop exercise participants.

**, Sample #2 for which its identity information was withheld from the tabletop exercise participants.

Table 2.2 Stakeholder meeting table top exercise sample identification results.

Sample	Data type	Group 1	Group 2	Group 3
Sample 1 (US_AL_L7256)	Petroleum biomarkers	AL	AL	AL
	GC-MS	AL	AL	AL
	IMS-MS	AL	AL	AL
Sample 2 (US_G_HI_L7291)	Petroleum biomarkers	G_EC	G_HI or G_EC	G_HI or G_EC
	GC-MS	G_HI	G_HI or G_EC	G_HI or G_EC
	IMS-MS	G_HI	G_HI	G_HI

Table 2.3 Table-top exercise general discussion questions and group-specific responses.

	Group 1	Group 2	Group 3
(i) How easy/difficult was it to identify the "unknown" samples using petroleum biomarker, GC-MS, or IMS-MS data? (on a scale from "1" = easy to "5" =difficult)			
Petroleum biomarker data	2	• Easier to see overall patterns • Not sufficient alone	• Unknown Sample 2 was difficult to distinguish
GC-MS data	4	• Easy when same well but different depths • Difficult because it is more information to compare • Not sufficient alone	• Unknown Sample 2 was difficult to distinguish
IMS-MS data	1	• Easy to recognize patterns • Dendrograms were very informative • Not sufficient alone	• Consistent grouping, known samples cluster better (additional confidence in the method)
(ii) How did the different data types compare in terms of their performance/quality/timeliness?			
Performance	• IMS-MS data were easier to inspect for patterns visually	• Traditional data (biomarkers and GC-MS) is regulatory accepted • IMS-MS is rapid, but unclear if there is easy access contract labs • IMS-MS cost and regulatory acceptance are unclear	• Alkanes data was least useful
Quality	• Data quality is hard to judge	• Data quality is hard to judge	• Data quality is hard to judge
Timeliness	• IMS-MS is most rapid • Multiple samples can be analyzed quickly	• Difficult to compare because IMS-MS is a new technique	• IMS-MS is most rapid • Rapid analysis enables analysis of technical replicates to increase confidence
(iii) What is needed to increase the use and application of IMS-MS data to facilitate the transition from research to practice?			
	• Combining IMS-MS with bioactivity data, not just GC-MS and biomarkers • Weathered sample analysis examples	• Broader representation of the IMS-MS library standards for identification of the components of petroleum samples • Database on weathered samples • Availability of the contract labs	• Commercial availability and standardization • Concurrence with traditional methods (GC-MS and biomarkers)
(iv) Overall, what is your impression of the IMS-MS and its proposed application to oil spill response?			
	• IMS-MS can be used as a quick scan because of its rapid sample processing and instrumental analysis	• Great potential, not as much training is needed to review the data outputs • Can be useful in emergency response situations • Availability is not clear, are there contract labs that use this technique?	• Great line of evidence in combination with other methods • Can be used for rapid screening of many samples then narrow down to fewer samples for further analyses • Useful for mixtures, analysis of degradation products

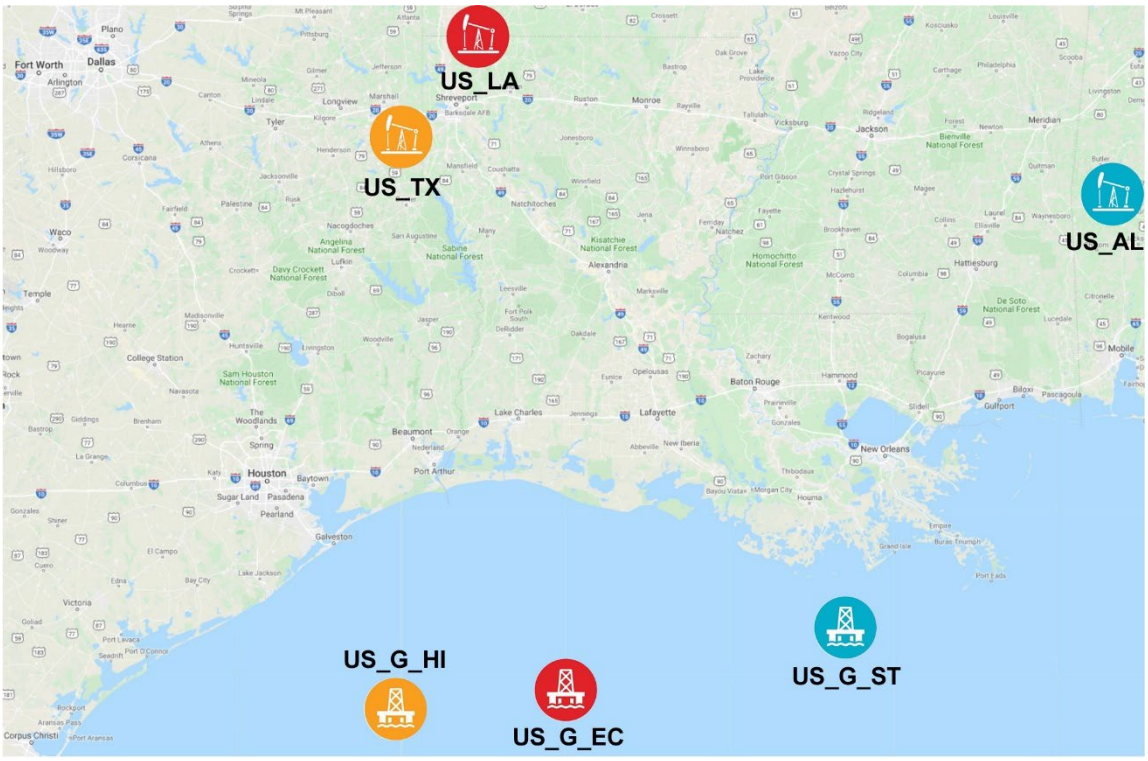


Figure 2.1 Crude oil samples (n = 19) were selected from 6 areas in 2 regions (onshore and offshore). Forensic fingerprinting analyses used GC-MS and IMS-MS technologies to evaluate the grouping of chemical profiles based on geographical extraction genesis. LA = Louisiana; TX = Texas; AL = Alabama; G = Gulf of Mexico; HI = High Island; EC = East Cameron; ST = South Timbalier.

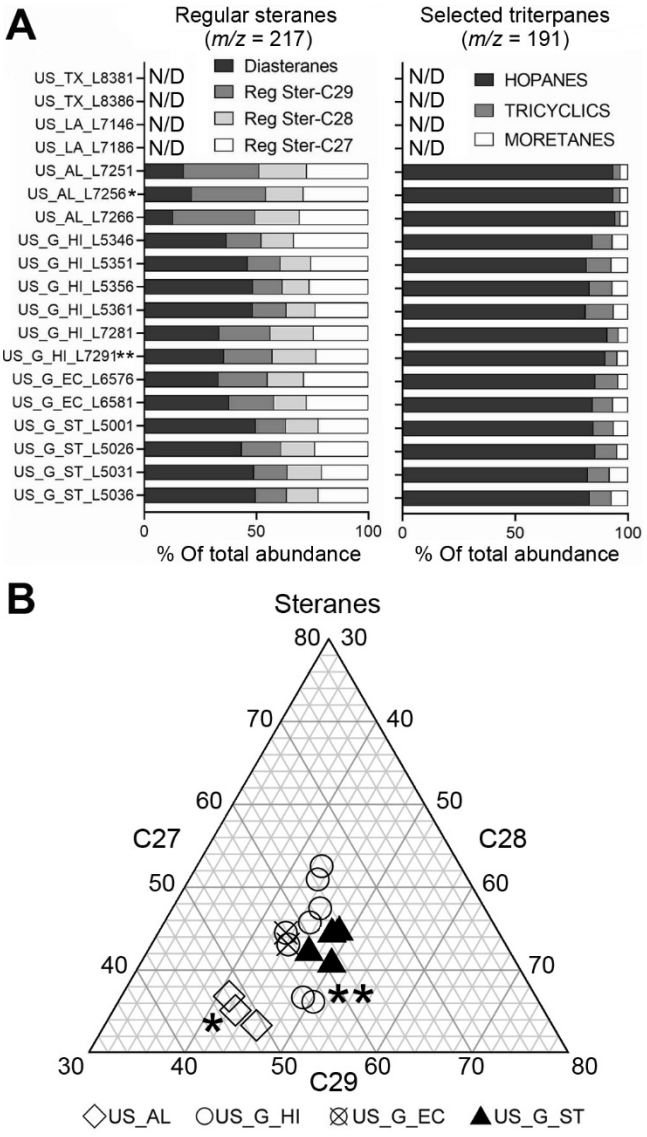


Figure 2.2 Gas chromatographic–mass spectrometric data processed for identification and quantification of biomarkers. (A) Distribution of regular sterane and triterpane biomarker ratios. (B) Ternary diagram of the relative distribution of C27, C28, and C29 regular steranes for crude oil samples. Asterisks identify blinded samples (*US_AL_L7256, **US_G_HI_L7291). N/D = not detected;

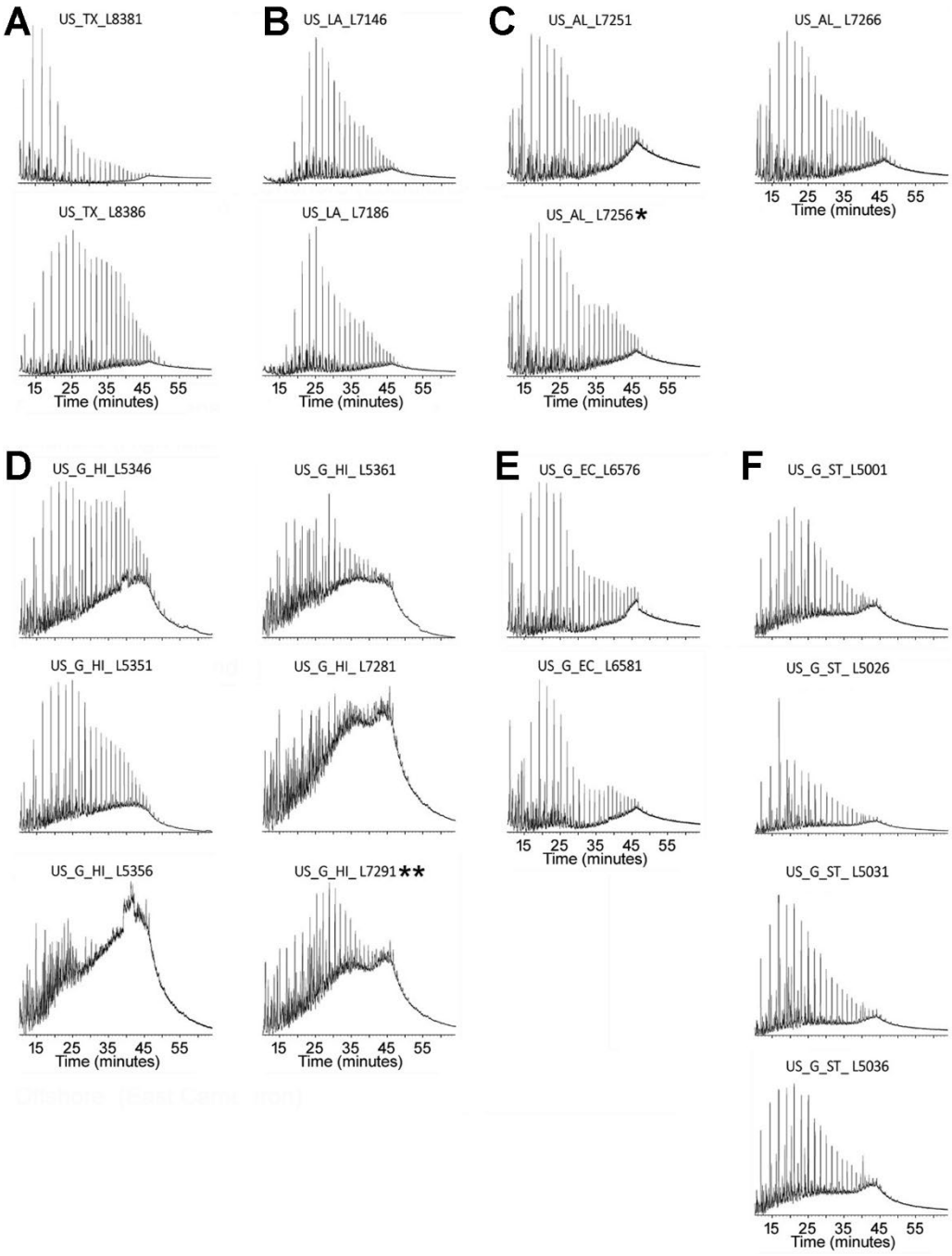


Figure 2.3 Gas chromatographic–mass spectrometric chromatograms (time vs abundance) for crude oil samples, a tool used for visual and qualitative analysis for forensic fingerprinting.

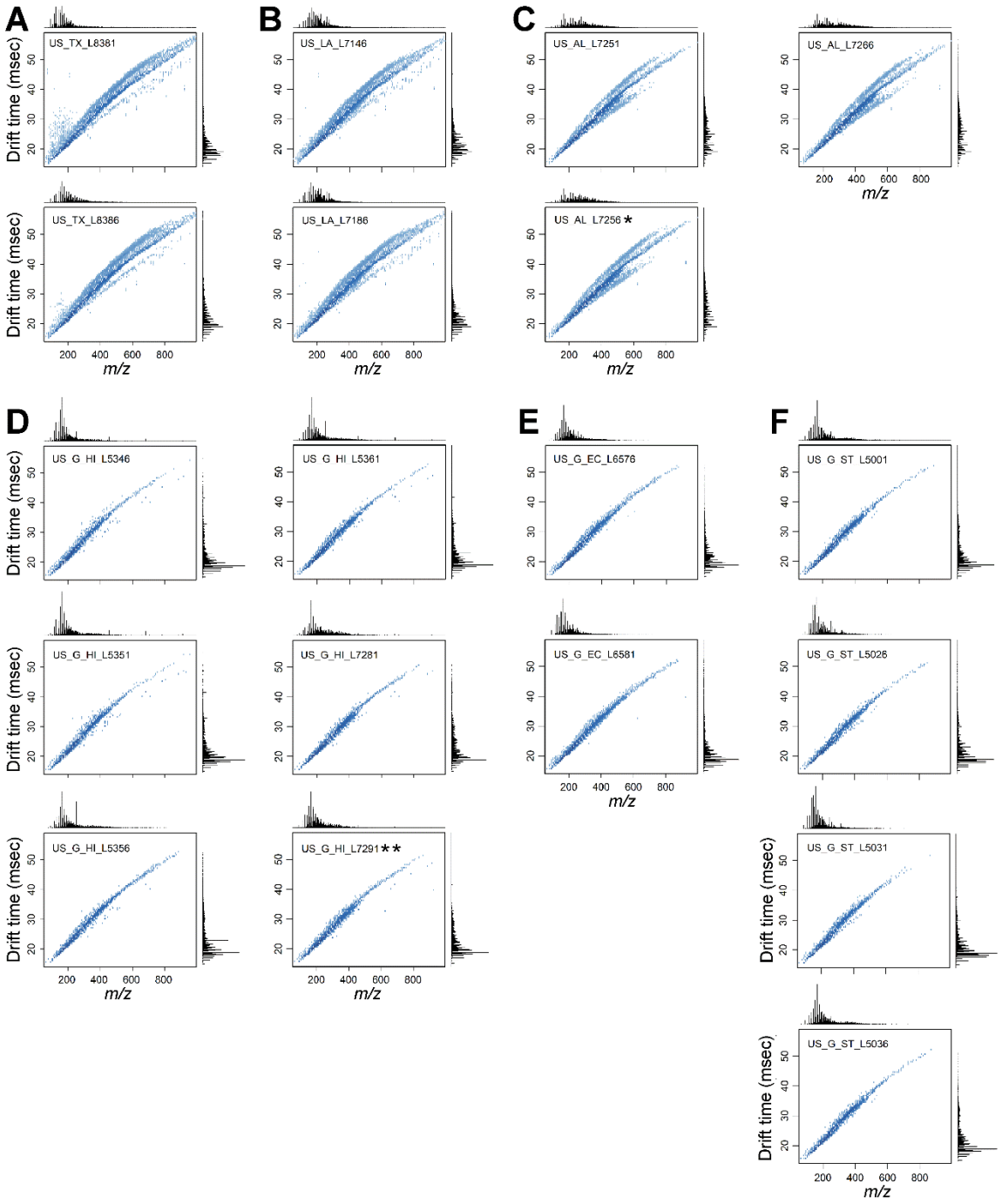


Figure 2.4 Ion mobility spectrometry–mass spectrometric spectra (*m/z* vs drift time, abundance is represented by color intensity) for crude oil samples, a tool used for visual and qualitative analysis for forensic fingerprinting. x-axes are *m/z*, and y-axes are drift time. Individual features are shown as dots. Density histograms of the features are shown at top (for *m/z*) and right (for drift time) on each plot.

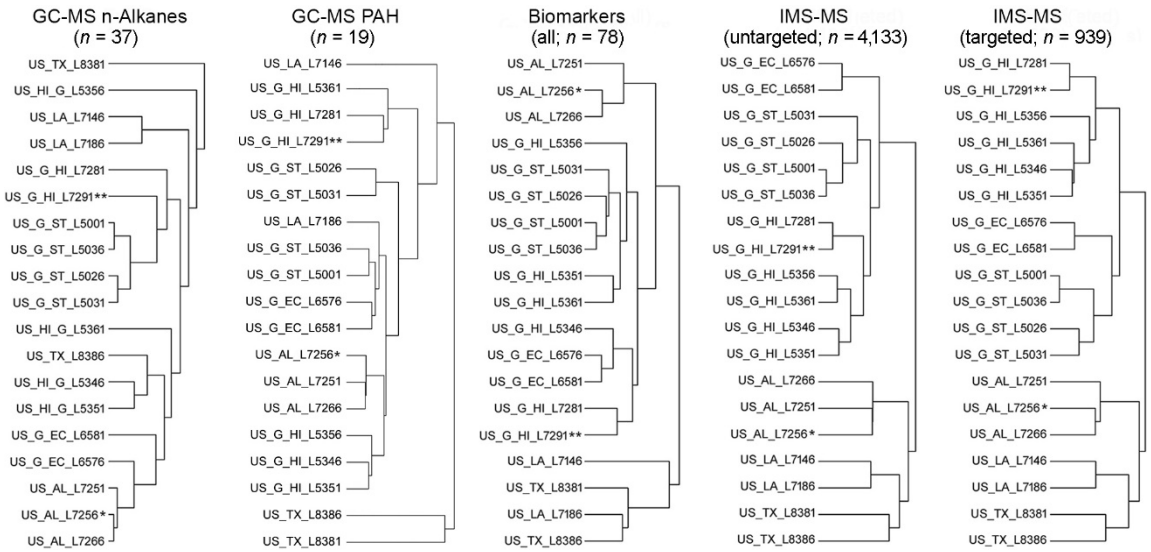


Figure 2.5 Unsupervised hierarchical clustering of crude oil samples from on- and offshore Gulf of Mexico region using data from different analysis methods. Each clustering diagram shown is labeled with the data set corresponding to Table 1. Number of features used for clustering is also identified. Gas chromatographic–mass spectrometric data sets were normalized and log-transformed. Other data were log-transformed. Asterisks identify blinded samples (*US_AL_L7256, **US_G_HI_L7291). See clustering diagrams for all comparisons included in Table 1 as Figure S2.13–37.

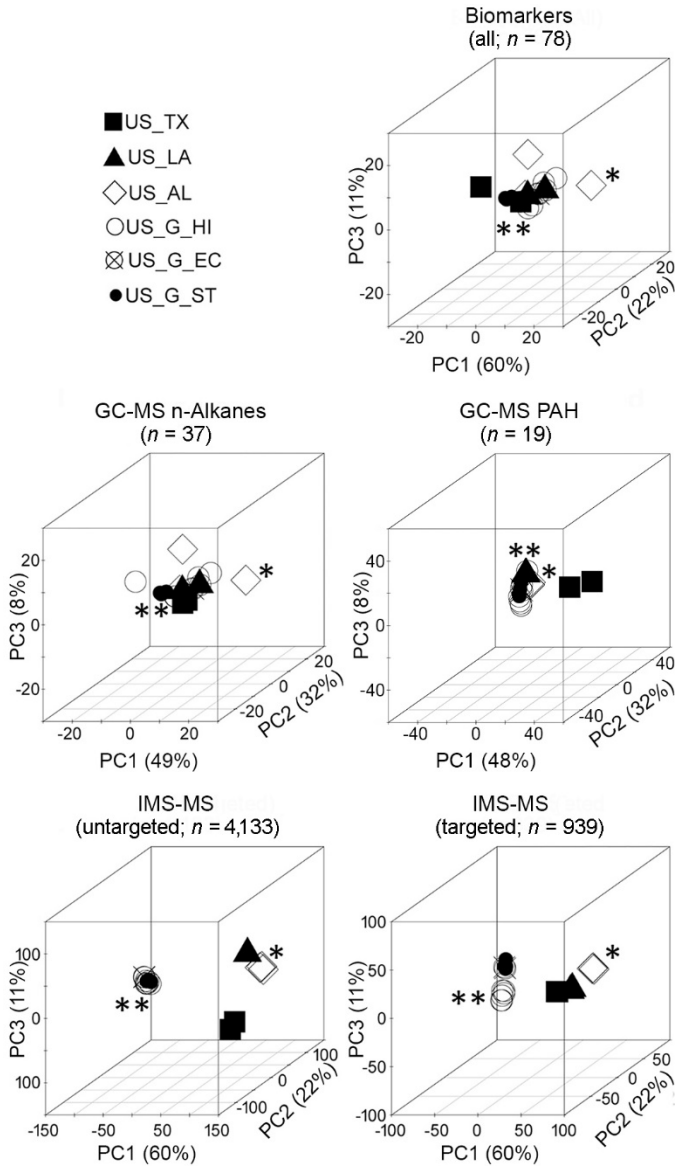


Figure 2.6 Principal component analysis grouping of crude oil samples (symbols indicate the area of origin as shown in the top left inset) from on- and offshore Gulf of Mexico region using data from different analysis methods. Each diagram shown is labeled with the data set corresponding to Table 1. Number of features used for analysis is also identified. Gas chromatographic–mass spectrometric data sets were normalized and \ log-transformed. Other data sets were log-transformed. Asterisks identify blinded samples (*US_AL_L7256, **US_G_HI_L7291).

3. CHARACTERIZATION OF COMPOSITIONAL VARIABILITY IN PETROLEUM SUBSTANCES*

3.1. Overview

In the process of registration of substances of Unknown or Variable Composition, Complex Reaction Products or Biological Materials (UVCBs), information sufficient to enable substance identification must be provided. Substance identification for UVCBs formed through petroleum refining is particularly challenging due to their chemical complexity, as well as variability in refining process conditions and composition of the feedstocks. This study aimed to characterize compositional variability of petroleum UVCBs both within and across product categories. We utilized ion mobility spectrometry (IMS)-MS as a technique to evaluate detailed chemical composition of independent production cycle-derived samples of 6 petroleum products from 3 manufacturing categories (heavy aromatic, hydrotreated light paraffinic, and hydrotreated heavy paraffinic). Atmospheric pressure photoionization and drift tube IMS-MS were used to identify structurally related compounds and quantified between- and within-product variability. In addition, we determined both individual molecules and hydrocarbon blocks that were most variable in samples from different production cycles. We found that

---

*Reprinted with permission from "Characterization of compositional variability in petroleum substances" by Alina Roman-Hubers, Alexandra Cordova, Arlean Rohde, Weihsueh Chiu, Thomas McDonald, Fred Wright, James Dodds, Erin Baker, and Ivan Rusyn, 2022. FUEL, 317, , Copyright 2022 by Elsevier.

detailed chemical compositional data on petroleum UVCBs obtained from IMS-MS can provide the information necessary for hazard and risk characterization in terms of quantifying the variability of the products in a manufacturing category, as well as in subsequent production cycles of the same product.

3.2. Introduction

Crude oil refining involves complex physical and chemical processes such as distillation, cracking, isomerization, reforming, alkylation and hydrodesulphurization, ultimately yielding petroleum products of certain performance characteristics that are subsequently used for a variety of applications (Kaiser 2017; Salvito et al. 2020). Because of the chemical complexity and variability of the oil feedstocks, as well as differences in the refining process conditions within and across manufacturing sites, it is expected that the types and quantities of hydrocarbons and other constituents present in downstream products may vary both within and between manufacturers, even for the same refining processes and products (CONCAWE 2012). This inherent compositional complexity and variability of petroleum substances, products that fall into the class known as substances of unknown, variable composition, complex reaction products, or biological materials (UVCBs), presents unique challenges for their registration and evaluation (Clark et al. 2013; ECHA 2017a). Current naming conventions and grouping of petroleum UVCBs into manufacturing categories is based on the information on their general composition such as carbon chain length and boiling point ranges, other physicochemical properties, performance characteristics, and proposed use(s) (Clark et al. 2013; Salvito et al. 2020). While Chemical Abstract Service (CAS) and European Inventory of Existing Commercial

Chemical Substances (EINECS) identifications have been assigned to petroleum products, they are further grouped into broad manufacturing categories for registration and regulatory evaluation (CONCAWE 2020b). The existing nomenclature for petroleum UVCBs, either for the individual product identifiers or for broad manufacturing categories, is deemed generally sufficient for the purpose of naming and identification of these products (Clark et al. 2013; ECHA 2017a; Salvito et al. 2020).

Once identified, petroleum substances must be registered following the laws and regulations of the jurisdiction where they are to be manufactured or used. The European Union REACH (Registration, Evaluation, Authorization and Restriction of Chemicals) regulation (Williams et al. 2009) specifies human health and the environment hazard data requirements that must be met before authorization is given for their use. Most petroleum UVCBs have considerable data gaps that need to be addressed in the process of registration either through additional testing or read-across to another substance, or other members in a group of substances, that have the requisite information for registration purpose (CONCAWE 2019c). Recent proposals for grouping of "similar" petroleum UVCBs into manufacturing categories, in part based on the "*broadly similar [chemical] composition*," have been questioned by the regulatory bodies as suitable for read-across, such as the European Chemicals Agency (ECHA). Concerns were raised about the strength of the justification for the proposed grouping and read-across (ECHA 2020). Recently, the European Commission has amended Annex XI of REACH clarifying that for the application of read-across/grouping, "*structural similarity for UVCB substances shall be established on the basis of similarities in the structures of the constituents (...) and*

variability in the concentration of these" (European Commission 2021). Indeed, the chemical variability in petroleum products is one well-appreciated concern, because petroleum substances "*are UVCBs and are manufactured to specifications based on performance characteristics rather than chemical composition, analysis of the same substance manufactured in the same location at different times could show a considerable variation in composition*" (CONCAWE 2019a).

To address the challenge of quantifying the variability of the petroleum UVCBs, a number of analytical approaches have been used to characterize their composition ranging from physicochemical analyses to detailed mass spectrometry (MS)-based methods (CONCAWE 2012; 2019a). Despite recent advances in petroleomics including novel high-resolution (HR) MS methods (Palacio Lozano et al. 2020), the ability to thoroughly assess the chemical composition of petroleum UVCBs, including the analysis of isomeric species, in the context of REACH has yet to be shown due to the similarity of the hydrocarbon components (*i.e.*, presence of isomeric species). In addition, the incomplete understanding of variability (*i.e.*, among samples from independent production cycles) represents a key barrier to the application of read-across between products in the same category, as any quantification of substance-to-substance similarity must be informed by within-substance variability. Accordingly, we set out to quantify both within- and between-product variability for representative petroleum UVCBs. We used both GC-MS IMS-MS techniques, because recent studies have demonstrated the utility of IMS-MS for determining the composition of petroleum UVCBs (Farenc et al. 2016; Ibrahim et al. 2016; Roman-Hubers et al. 2021a; Roman-Hubers et al. 2021b; Ruger et al. 2021). We identified

structurally related hydrocarbon and heteroatom compounds, examined hydrocarbon blocks, and characterized within-product variability.

3.3. Materials and Methods

3.3.1. Samples of petroleum products

A total of six refined petroleum products (Table 3.1) were used in this study. For evaluation, we selected three products from three broad manufacturing categories of "Solvent naphtha, heavy aromatic products" (marked as AR) and "Petroleum distillates, hydrotreated [light or heavy] paraffinic" (marked as BO). Sample selection was meant to be representative of a wide range of expected chemical complexity of petroleum UVCBs. For each product, samples were obtained from 2-3 independent production cycles at the same refinery (samples were collected 2-3 months apart), resulting in a total of 16 samples (Table S3.1). For GC-MS analyses, samples were weighed and dissolved in dichloromethane (CAS no. 75-09-2, catalog no. 34856; Sigma-Aldrich, St. Louis, MO) to a final concentration of 1 mg/mL. For the IMS-MS analyses, 1 mg of each sample was first dissolved in 9 mL of 1:1 (v/v) mixture of toluene (CAS no. 108-88-3, catalog no. 34866; Sigma-Aldrich) and methanol (CAS no. 67-56-1, catalog no. 34860; Sigma-Aldrich). Next, 25 μL of the solution was mixed with 300 μL of the same mixture of toluene and methanol and injected directly.

3.3.2. GC-MS instrumental analysis and data processing

The modified United States Environmental Protection Agency (US EPA) method 8270 was carried out in full scan analysis mode using an Agilent 7890 GC (Agilent Technologies, Santa Clara, CA) interfaced with a Hewlett-Packard (HP) 5976 MS.

Additionally, a HP-5ms Ultra Inert Column (30 m × 0.25 μm × 0.25 mm; catalog no. G3900-63001; Agilent Technologies) was used to chromatographically separate the petroleum hydrocarbons. Instrumental operating conditions were as follows: mass range 40 to 500m/z, splitless injector, injection volume of 2 μL, column flow 1 mL/min, helium carrier gas. Initial temperature of the injection port was held at 250°C. The oven was initially set to 50°C with a hold time of 4 min; then, the oven was programmed at a rate of 6°C/min until it reached the final holding temperature of 300°C with a final hold time of 20 min. Individual full-scan total ion chromatograms for each sample were processed using ChemStation Data Analysis Software (Agilent). Raw data consisted of 10,127 scans at 1 atomic mass unit (amu) bins from 40 to 500 amu. Data for each amu bin across all scans was averaged and the final data matrix consisted of an average abundance value for each amu bin for one sample. The data for 16 samples were combined into a two-dimensional data matrix of mass range versus average fragment ion intensities. See Table S3.**2** for the resulting GC-MS data matrix and Figure S3.1 for the GC-MC chromatograms for each sample.

3.3.3. IMS-MS instrumental analysis and data processing

For the IMS-MS analyses, we utilized an Agilent Technologies G6560A platform coupling drift tube IMS (resolving power (RP) ≈ 60) and a quadrupole time-of-flight (QTOF) mass spectrometer (RP ≈ 25 000). In all experiments, the drift tube was filled with nitrogen gas and the samples were ionized with an atmospheric pressure photoionization (APPI) source (model G1917C; Agilent Technologies). The instrument was calibrated prior to running samples according to the Agilent protocol for 50-1,700 *m/z*

range, using the atmospheric pressure chemical ionization (APCI)-L Low Concentration tuning mix solution (part #G1969-85010, Agilent Technologies). The petroleum samples (200 μL) were then infused directly at a flow rate of 50 μL/min, and analysis included three technical replicates for each sample. Instrumental and source parameters were as follows: APPI positive ion mode, sample analysis time 1.5 min; source parameters: gas temperature 325 °C, vaporizer 350 °C, drying gas 10 L/min, nebulizer 30 psi, VCap 3000, fragment 400 V, 110 RF Vpp 750. The following acquisition parameters were defined for each instrumental run: mass range 50 to 1700 *m/z*, frame rate 1 frames/s, IM transient rate 18 transients/frame, maximum drift time 60 ms, time-of-flight transient rate 600 transients/IM transient, trap fill time 20 000 μs and trap release time 300 μs. QTOF parameters were as follows: firmware Ver 18.723, Rough Vac 2.71 torr, Quad Vac 3.68 × 10−5 torr, TOF Vac 3.47 × 10−7 torr, drift tube pressure 3.940 torr, trap funnel pressure 3.790 torr, chamber voltage 5.96 μA, and capillary voltage 0.076 μA. Data were obtained using the MassHunter Acquisition software (Agilent; ver. 08.00). Each sample was analyzed in triplicate in three independent experimental batches using the instrument and setting as detailed above. Two of these experiments were conducted at Texas A&M University on separate days (about 1 month apart) by different individuals. One of the experiments was conducted at North Carolina State University. These replication studies were using the same model IMS-MS instrument and experimental conditions, but were conducted by three different individuals.

IMS-MS raw data files from each instrumental run were processed using MassHunter Browser Acquisition Data software (Agilent Technologies, ver. 08.00) to

derive nitrogen gas-filled drift tube collisional cross section ($^{DT}CCS_{N2}$) values for all detected features (Stow et al. 2017). In this manuscript, a feature is defined as a potential molecule's isotopic envelope and by having both MS and IMS dimensions, all isotopes must occur at the same IMS drift time. Next, data files for all samples and their respective technical replicates (16 samples × 3 technical replicates = 48 files) were uploaded to Agilent MassProfiler software (Agilent Technologies, B.08.00) for feature alignment based on drift time (±5.0%) and mass (±15ppm+5mDa). Finally, aligned raw data matrices for each experimental batch (Table S3.3) were filtered to select features with abundance >5,000 in two out of three technical replicates for each sample. These filtered data (Table S3.4) include information on the constituents present in high abundance that would be of most relevance with regards to hazard evaluation of petroleum UVCBs (McKee et al. 2005). The filtering parameters were selected based on the general consideration of the presence of ^{13}C isotopic partner for individual features and previous data analyses (Roman-Hubers et al. 2021a) that showed erosion in confidence for molecular formulae assignments for the features of low abundance; however, alternative thresholds may be selected using the datasets provided in Table S3.3.

After alignment and filtering as detailed above, the data (Table S3.4), including technical replicates (n=3), was used for feature identification using an IMS-MS data processing workflow detailed elsewhere (Roman-Hubers et al. 2021a). Briefly, each feature was cross-referenced to a $^{DT}CCS_{N2}$ standard library containing a number of hydrocarbon standards (Baker 2021). Features were deemed matching to a molecule in the database at a $^{DT}CCS_{N2}$ tolerance of ±1% and an *m/z* tolerance of ±5 ppm and ±2 mDa.

Then, the Kendrick Mass Defect (KMD) was calculated for each feature using base units of CH_2 (14.01565) and H (1.00783) to identify features that fall into homologous series (±1.00 parts per thousand, ppt). Next, the elemental composition was assigned to each feature if it was in homologous series using the $^{DT}CCS_{N2}$ library matched features as reference points, as well as based on the KMD-H analyses as detailed in (Roman-Hubers et al. 2021a) (Table S3.**5**). Carbon chain length and double bond equivalency (DBE) of each feature were calculated from the elemental composition (McLafferty F.W. and F. 1993). Based on the elemental formula and other properties, each feature was assigned a carbon chain length and hydrocarbon class.

3.3.4. Data analysis

We reasoned that quantifying overall similarity among samples would be instructive to illustrate the informativeness of the data. Thus, the GC-MS and IMS-MS data matrices of all samples (Table S3.2-4) were used for hierarchical clustering (Everitt 1980) based on Spearman correlation and average linkage using the *hclust* package in R Studio (ver. R-4.1.0). The correlation among samples was then visualized in a dendrogram. In order to assess the similarity between clusters, the Fowlkes-Mallows (FM) index (Fowlkes and Mallows 1983) was used to evaluate the concordance of experimental data-derived clustering to that of the pre-defined manufacturing categories, products, independent production cycles, and technical replicates of each sample. The technical replicates and independent production cycles were considered as separate instances of the same substance, and the FM index was compared between the pre-defined categories and the hierarchical clustering having a number of clusters equal to the number of

manufacturing categories, using the *cuttree* command on the clustered tree. FM index values can range from 0 (no correspondence) to 1 (perfect correspondence). Principal components analysis was carried out to evaluate similarity between the products and samples using the *prcomp* and *ggplo*t packages in R Studio (4.1.0) and based on characterized features, carbon chain length, hydrocarbon class and heteroatom species. For analysis at the individual ion level (*i.e.*, full dataset of 55,466 features), the differences in abundance of each feature in samples from the independent production cycles were quantified as the maximum of the absolute value (if at least 3 samples were available) of the fold change difference when comparing across all pairs of samples from independent production cycles. For each feature, a p-value for variability was assessed by a one-way analysis of variance, using production cycle as a factor. Correction for multiple comparisons across features was performed using the Benjamini-Hochberg *q*-value computed in R using *p.adjust* (Table S3.**6)**. For the analysis at the level of carbon chain length, hydrocarbon class and heteroatom profile, the variability in chemical composition among samples of independent production cycles was evaluated based on the relative abundance of the molecules in each aggregated set of features using two-way ANOVA with Sidak's multiple comparison test (GraphPad Prism 9.0, San Diego, CA) followed by the Bonferroni correction (Dunn).

3.4. Results

This study evaluated 16 samples of 6 oil refining-derived products that fall into three broad manufacturing categories of petroleum UVCBs (Table 3.1). We began by analyzing samples using conventional GC-MS technique. Figure 3.1A shows

superimposed GC-MS full-scan chromatograms for representative samples of each product (see Table S3.1 for the similar chromatograms of each sample). These data clearly demonstrate the difference among samples from diverse manufacturing categories and the "cuts" of hydrocarbons varied considerably among samples as evidenced by the retention time differences. The "solvent naphtha (petroleum), heavy aromatic" products AR150 and AR200 were readily separated by GC-MS. The "distillates (petroleum), hydrotreated" light (BO60), or heavy (BO100, BO220 and BO600) products were more complex as they yielded characteristic unresolved complex mixture (UCM) "humps" on the GC-MS chromatograms. To visualize the similarity among products in the GC-MS data, we used the data matrix of averaged intensities for each of the amu bins (from 40 to 500 amu) to conduct unsupervised hierarchical clustering analysis. Figure 3.1B shows that samples from independent production cycles of the same product clustered together, except for one BO220 sample (production cycle 1). Moreover, solvent naphtha samples and the hydrotreated paraffinic distillate samples also formed distinct groups commensurate with their manufacturing category and CAS# groupings (Table 3.1). Samples of product BO60 from independent production cycles were most dissimilar to each other, yet they still clustered into their own group. Based on GC-MS data, the concordance in the clustering of the samples, as compared to pre-determined manufacturing category assignments for each sample, was modest (FM index of 0.49).

Samples were next analyzed using the IMS-MS platform. Figure 3.2 shows representative two-dimensional nested spectra for representative samples of each product where they are plotted by *m/z* (x-axis) and drift time (y-axis, parameter used to calculate

$^{DT}CCS_{N2}$) with feature abundance represented by color intensity. Figure S3.2 shows IMS-MS nested spectra for each sample analyzed. These plots illustrate the differences in both complexity (total number of features) and changes in *m/z* and structural sizes (IMS) of the individual constituents in the samples. For example, samples of solvent naphtha (petroleum) heavy aromatic products AR150 and AR200 contained compounds in the mass range of 50-400 *m/z*. The hydrotreated distillates light (BO60) product contained lower *m/z* range species as compared to the hydrotreated distillates heavy (BO100, BO220 and BO600) products that spanned a mass range of up to 700 *m/z*.

Unsupervised hierarchical clustering utilizing *m/z*, $^{DT}CCS_{N2}$ and feature abundance data from the IMS-MS analyses was then used to compare composition similarity among the samples. For these analyses, both the full data matrix (Table S3.3) and filtered (i.e., most abundant features) data was assessed (Table S3.4). Figure 3.3 illustrates the full and filtered abundance dendrograms where the technical replicates were averaged for the analyses. When the full IMS-MS data was used (Table S3.3A), samples from independent production cycles of the same manufacturing category clustered together resulting in three main clusters. For example, samples of BO60 product clustered closer to the AR150 and AR200 products and not with the other products (BO100, BO220 and BO600). Additionally, the concordance in sample clustering for pre-determined manufacturing category assignments was excellent (FM index of 1.0). When the IMS-MS datasets were filtered for only highest abundance features (Table S3.4A), similar clustering was observed, and the FM index for this analysis was also 1.0. Furthermore, when technical replicate samples were included in these analyses, similar results were

obtained (Table S3.3). Specifically, technical replicates of each sample clustered together, and then with the samples from different production cycles for each product, and finally with other products within a manufacturing category.

Our next assessment was to determine how well petroleum UVCBs group using IMS-MS data obtained in independent experiments by distinct operators and in a different laboratory. For these studies, the samples were analyzed at Texas A&M on the same instrument but by a different operator and then at North Carolina State University by another operator and instrument, but the same model of IMS-MS platform (G6560) and an identical experimental protocol. In all cases, samples were prepared independently from the neat stocks of each product (see Methods) before each instrumental analysis. Abundance-filtered data (Table S3.4), where technical replicates were averaged, were used for the following comparisons. Irrespective of the laboratory or operator, strong correlation between samples from independent production cycles was evident as products clustered within their manufacturing category and CAS# (Figure 3.4 and Table S3.7). The FM index values for clustering were 1.0 for two experiments (Figures 3.4A-B) and 0.86 for the third one (Figure 3.4C). These results indicate high reproducibility of the IMS-MS technique for the analysis of similarities in samples of complex-composition petroleum UVCBs.

While clustering using multidimensional data from untargeted IMS-MS is useful to establish the overall similarity of the samples for the purpose of substance identification (Roman-Hubers et al. 2021b), this information may not be adequate for product registration because it does not provide sufficient detail on the chemical composition of

each sample. To address this challenge, we identified structurally related compounds (Roman-Hubers et al. 2021a) in analyzed petroleum products to obtain molecular formula assignments to the high abundance features. Because each of three independent IMS-MS experiments (Figure 3.4) yielded similar clustering of the samples, filtered IMS-MS data from one of the experiments was used herein (Table S3.4A). Molecular formulas for each feature in the 16 samples are provided in Table S3.5. Similar to our previous findings of analysis of refined products or crude oils (Roman-Hubers et al. 2021a), we were able to assign molecular formulas to 93% of the high abundance features across all samples.

Because the composition of petroleum UVCBs is typically presented using the hydrocarbon block method which groups closely related compounds by their carbon chain length and hydrocarbon class (CONCAWE 1996; 2019b), we used the assigned molecular formulas and other information from the KMD analysis (*i.e.*, homologous series and double bond equivalence) to aggregate the data into hydrocarbon blocks (Table S3.8). We also determined whether any of the identified molecules were heteroatoms (Table S3.9). With these data, we performed principal component analysis to visualize the similarities between samples of independent production cycles, as well as differences among products across manufacturing categories (Figure 3.5). When all high abundance features with assigned molecular formulas (n=1,417, Figure 3.5A) or carbon chain length (Figure 3.5B) were used for the principal component analysis, four groups were discernable in the first two principal components. Group 1 and Group 2 distinguished between two heavy aromatic products (AR150 and AR200), while Group 3 separated the light (BO60) and Group 4 contained the heavy (BO100, BO220 and BO600) hydrotreated paraffinic

distillate products. Interestingly, the latter group appeared more homogenous even though it contained samples from three different products. Additionally, samples from independent production cycles were closely aligned to each other. The molecular formula-level data showed tighter grouping between samples of the same product, while the data on hydrocarbon blocks (Figure 3.5C) or heteroatom profiles (Figure 3.5D) allowed fewer distinctions among product groups, there was wider separation between samples from production cycles.

To further evaluate the variability in petroleum UVCBs, we analyzed the relative abundance of the molecules in hydrocarbon blocks or heteroatoms between samples from independent production cycles of the same product (Table S3.8-9). Figure 3.6 shows an example of this analysis for product BO220. Figures 3.6A-C show the relative abundance of each hydrocarbon block, as well as total abundance for each carbon chain length and hydrocarbon class. It is evident that while the overall ranges in carbon chain length and hydrocarbon classes were largely concordant, the abundances of the constituents in each hydrocarbon block varied. Significant differences were observed in most highly abundant hydrocarbon blocks (Figures 3.6D-E). In addition, the relative proportion of O_1-containing heteroatoms was also significantly different between production cycles (Figure3 3.6F).

Similar analyses were performed for each product and the quantitation of the variability is presented in Figure 3.7. For the carbon chain length data (Figure 3.7A), product BO220 was most variable in terms of the number and range of molecules that were significantly different between production cycles. Even though product BO600 was equally complex in terms of the overall range of hydrocarbons, only C_{35}-containing

molecules varied significantly between production cycles. Products AR150 and AR220 showed variability in about one-third of the hydrocarbon blocks. Product BO60 showed no variability, and product BO100 showed variability in only a few blocks; however, there were only two samples available for those products and therefore limited variability should be interpreted with caution. Similar findings were observed with the hydrocarbon class data (Figure 3.7B). Most production cycle-associated variability was found in mono-, di- and tri-aromatic compounds. For the heteroatom data (Figure 3.7C), only products AR150 and BO220 showed variability with O_1-containing molecules, these were the most abundant and variable heteroatoms.

Even though the analysis of within-product variability based on the hydrocarbon block method has broad utility, such data are deemed insufficient in terms of satisfying the regulatory need for "*detailed chemical characterization*" of petroleum UVCBs for product registration purposes. Therefore, we also used the data on high abundance features to characterize the variability between samples of each product that were derived from independent production cycle. For this, we examined (i) the degree to which the individual features varied between samples from different production cycles (*i.e.*, p-values for each molecule), and (ii) the average relative abundance of each feature in a product across production cycles (Table S3.10). Figure 3.8 shows the results of this analysis for each product. The figure shows that the feature abundance threshold set during data processing was concordant with the REACH Regulation threshold of 0.1% abundance for constituents of concern in complex substances (ECHA 2017b). It is evident that there are hundreds of constituents in each examined product that are present in quantities above 0.1%. These

included expected abundant amounts (5-10%) of naphthalene and related mono- and di-aromatic hydrocarbons (Table S3.11). However, few constituents were significantly different between samples (using a cutoff based on the Bonferroni-corrected p-value which was a false discovery rate of 5% corrected for the total number of features) of the same product from independent production cycles. We found no constituents that are both significantly different and reasonably abundant in products BO60, BO100 and BO600, even though these products spanned the degree of complexity of the entire dataset in terms of the number of high abundance features. Product AR200 had the largest number of constituents identified as variable and abundant. To the contrary of the results with a hydrocarbon block method data analysis (Figures 3.6-7), product BO220 only had 3 constituents above the variability and abundance thresholds, even though the total number of constituents with suggestive significance was large. Product AR150 had only one constituent above the thresholds. Table 3.2 lists the constituents for each product that were identified as above the thresholds in Figure 3.8. A number of the constituents identified in these analyses are currently listed by ECHA as Annex III substances, which are substances predicted to likely present health or environmental hazard (ECHA 2016). One constituent in product AR200, anthracene, a feature whose identity was identified using IMS-MS data from a chemical standard, is identified by ECHA as a substance of very high concern.

3.5. Discussion

The analytical chemistry challenges in petroleomics are many (Marshall and Rodgers 2004; 2008), and the potential solutions range from well-established physicochemical and analytical methods (CONCAWE 2012; 2019a) to novel high-

resolution mass spectrometry techniques (Cho et al. 2015; Islam et al. 2012; Palacio Lozano et al. 2020; Santos et al. 2015; Terra et al. 2014). However, there appears to be a growing chasm between the research-driven advances in high-resolution mass spectrometry for petroleomics, and the needs of the practitioners in the industry and regulatory agencies. "*Sufficient*" characterization of highly complex petroleum UVCBs as products allowed into commerce and trade at various economic areas, such as the European Union where REACH defines data requirements (Williams et al. 2009), is a pressing regulatory need. As recently as 10 years ago, it was noted in a report by a major trade association of the petroleum refiners in Europe that conventional MS-derived "*data obtained by direct analysis of a petroleum UVCB substance, in which all constituents are ionized and fragmented simultaneously, would be too complex to allow meaningful interpretation*" (CONCAWE 2012). Indeed, regulatory submissions of petroleum UVCBs do not typically use conventional MS-based data, or more contemporary high-resolution mass spectrometry data, but rather include information that defines chemical composition broadly, for example into hydrocarbon blocks (CONCAWE 1996; 2019b). The typical substance identity information provided to the regulators such as ECHA consists of the manufacturing process description, various physicochemical data (boiling point and carbon chain length ranges, etc.), relative proportions of constituents in major hydrocarbon classes (saturates, aromatics, resins, asphaltenes, etc.), and relative content of various polycyclic aromatic compounds (by the number of aromatic rings). Invariably, regulators express dissatisfaction that the individual chemical constituents, their structural features, and quantitative metrics for the intrinsic variability of the products that are being

registered are not attainable using the analytical methods on which the industry is relying heavily. For example, in a recent decision from ECHA on a testing proposal for grouping of substances in the "Residual aromatic extracts" manufacturing category, the agency concluded that chemical similarity between products has not been established for the purpose of registration (*i.e.*, read-across), because "*no qualitative or quantitative comparative assessment of the compositions of the different category members*" has been presented (ECHA 2020). This representative ECHA decision further noted that because of the intrinsic compositional variability of petroleum UVCBs, detailed information in support of the "*chemical similarity*" argument would need to include (i) detailed data on the composition of the test sample(s) (both individual constituents and "major hydrocarbon classes"), as well as (ii) data on intrinsic chemical variability among products in a category (ECHA 2020), these requirements were recently added to Annex XI of REACH (European Commission 2021).

A number of recently developed multidimensional high-resolution mass spectrometry techniques, including IMS-MS, when applied to the analysis of petroleum samples, demonstrated excellent molecular resolution and the ability to characterize high molecular weight hydrocarbons, including isomeric species (Fernandez-Lima et al. 2009; Grimm et al. 2017; Guillemant et al. 2019; Hsu et al. 2011; Marshall and Rodgers 2004; Niyonsaba et al. 2019; Palacio Lozano et al. 2019b; Ponthus and Riches 2013; Roman-Hubers et al. 2021b; Santos et al. 2015). In addition, a number of chemometric methods have been proposed in conjunction with high-resolution mass spectrometry data on oil and petroleum products for fingerprinting and source identification (de Carvalho Rocha et al.

2017; Niyonsaba et al. 2019; Onel et al. 2019; Palacio Lozano et al. 2020), as well as the capability of the molecular and structural identity of chemical constituents and hydrocarbon blocks (Gabelica et al. 2019; Koch et al. 2007; Roman-Hubers et al. 2021a). Therefore, we reasoned that the opportunity exists to demonstrate the value of multidimensional high-resolution mass spectrometry petroleomics as a solution to current challenges in chemical characterization of petroleum UVCBs for regulatory decision-making purpose. To this effect, this study aimed to demonstrate how one of the multidimensional high-resolution mass spectrometry petroleomics techniques, IMS-MS, can be used to address the regulatory challenges by (i) providing qualitative and quantitative information on the composition of representative complex petroleum substances, and (ii) using this information to characterize the variability of the constituents in the substances manufactured in different production cycles or those grouped into the same broad category. Our choice of APPI ionization in positive mode with IMS-MS as an analytical technique was informed by prior studies demonstrating improved resolution of isomeric aromatic species in petroleum samples (Borsdorf et al. 2006; Roman-Hubers et al. 2021a). Specifically, we aimed to take advantage of the IMS-MS technique-derived data on the differences in drift time among various hydrocarbons of the same atomic composition (*i.e.*, isomeric species), rather than focus on increasing the resolution in the *m/z* dimension, a common goal in petroleomics studies afforded by ultrahigh resolution Fourier transform ion cyclotron resonance (FTICR) MS (Marshall et al. 2010) and other high-resolution mass spectrometry techniques (Niyonsaba et al. 2019). Recent studies demonstrated the utility of IMS-MS for determining the chemical composition of

petroleum substances and crude oils (Farenc et al. 2016; Ibrahim et al. 2016; Roman-Hubers et al. 2021b; Ruger et al. 2021) and we proposed a chemometric method for deducing the chemical compositional information for both refined products and crude oils that uses $^{DT}CCS_{N2}$ information to increase confidence in the evaluation of the chemical composition of the features in homologous series (Roman-Hubers et al. 2021a). Overall, we hypothesized that high resolution untargeted IMS-MS analysis, in conjunction with a petroleomics data processing workflow and chemometric evaluation, would enable detailed characterization of the most abundant ionizable molecules in petroleum UVCBs, providing quantitative data on substance-to-substance variation that will inform overall hazard assessment. To test this hypothesis, we evaluated both a range of petroleum products, and samples from independent production cycles of the same product.

Overall, we highlight four major advances afforded by this study. First, we demonstrate how IMS-MS data can be used to evaluate broad similarity among substances while also identifying the degree of variability within a class or between production batches of the same substance. By comparing and contrasting the IMS-MS data to that from GC-MS, we confirm advantages in both resolving power, and coverage of the high molecular weight compounds. GC-MS is used widely to characterize the composition of various fuels and to classify and group the fuels (de Carvalho Rocha et al. 2017). In addition, GC×GC-flame ionization detection technique (CONCAWE 2019b; Frysinger et al. 2003; Gaines et al. 1999; Van De Weghe et al. 2006) is also commonly used for petroleum analyses to derive “hydrocarbon blocks” for substance identification purposes (Bierkens and Geerts 2014; CONCAWE 1996). It was previously shown that the

multidimensional data from these techniques can be used for fingerprinting of oils or grouping petroleum UVCBs, but that IMS-MS data typically affords greater classification and fingerprinting accuracy (Onel et al. 2019; Roman-Hubers et al. 2021b). In this study, we found a similar pattern, with IMS-MS data superior to that from GC-MS for grouping and classification of the samples.

Second, this study goes farther than grouping and classification as we were able to assign confident molecular formulas to most (on average 93% across all samples) of the high abundance features from IMS-MS data. To achieve this, we selected only the highest quality abundant features at the expense of focusing on a relatively small fraction (~2%) of all detected features. While the process of dimensionality reduction may seem counter to the desire to provide as detailed chemical characterization of the UVCBs as possible, the following considerations support our approach: (i) the confidence in molecular formula assignments for the features beyond those with highest abundance erodes rapidly (Roman-Hubers et al. 2021a), (ii) even though IMS-MS is able to resolve tens of thousands of features in petroleum UVCBs, numerous molecules are still undetected either due to ion suppression or instrument sensitivity (Hawkes et al. 2020; Santos et al. 2015), and (iii) if these chemical composition data are to be used for regulatory decisions, it is acknowledged that the priority shall be given to the highest abundance constituents in complex substances. For example, according to Articles 7(2) and 33 of REACH Regulation (ECHA 2017b), the abundance threshold of 0.1% (w/w) is set (for the purposes of either notification of substances in articles, or communication of information on substances in articles) for constituents that are classified as substances of very high

concern. This implies that the focus on the highest abundance features when analyzing detailed chemical composition of petroleum UVCBs would be responsive to REACH Regulation requirements, because other molecules in each sample are likely present at amounts far below the 0.1% threshold.

Third, a very important consideration for the use of an analytical method for regulatory decision-making is its accessibility and reproducibility. Both GC-MS and GC×GC with flame ionization detection are used to generate data for regulatory submissions because these methods have been standardized (ASTM International 2011; US EPA 2014). In this regard, commercialization of the drift tube IMS-MS made these instruments available in a standard configuration leading to a growing number of publications demonstrating their use for petroleomics (Roman-Hubers et al. 2021a; Roman-Hubers et al. 2021b; Santos et al. 2015). In addition, studies of reproducibility of IMS-MS-derived experimental parameters such as standardized drift tube, nitrogen CCS values ($^{DT}CCS_{N2}$) were conducted using hundreds of molecules across multiple laboratories and illustrated the potential of this technique for providing confident molecular identifiers for a broad range of discovery-based analyses (Baker 2021; Stow et al. 2017). This study, while not a formal cross-laboratory standardization analysis, does demonstrate that samples can be confidently compared across operators in the same laboratory and across laboratories. Therefore, this technique and approach have promise for wider application as they are based on a commercially-available instrument and also a fairly rapid analysis based on gas phase separations and direct injection that does not require extensive sample preparation.

Finally, because of the ability to deduce molecular identifications for hundreds of molecules in complex petroleum UVCBs, a number of existing challenges with chemical characterization of petroleum UVCBs for hazard assessment are potentially resolved. Specifically, it is possible to identify constituents and determine their abundance for consideration as potential substances of concern. Because the hydrocarbon block method (Bierkens and Geerts 2014) is widely used for the characterization of human health and environmental hazards of petroleum UVCBs, the IMS-MS data with molecular identifiers can be used to construct data matrices similar to those generated in GC×GC with flame ionization detection, but where the identity of the constituents in each block are known. In addition, the variability between independent production cycles and among samples in the same product category can be quantitatively characterized; if it is found that samples are significantly variable, it is now possible to determine whether such variability may impact potential hazardous properties of the entire substance and reduce uncertainty in grouping.

One limitation of this study, similar to other analytical studies of petroleum UVCBs, is that the complete chemical characterization of petroleum UVCBs is unattainable. The extent of the molecular resolution depends on the type of ionization and detection methods and instruments, as well as sample processing and other factors (Palacio Lozano et al. 2019a). For example, the APPI ionization used herein, albeit a preferred method for characterization of nonpolar petroleum fractions (Kauppila et al. 2002; Purcell et al. 2007a), is not applicable to the analysis of paraffins. Still, our method is suitable for evaluation of polycyclic aromatic compounds, which include polycyclic aromatic

hydrocarbons and heteroatoms, substances that have been associated with carcinogenic activity (Ayala-Cabrera et al. 2021; McKee and White 2014). We also note that other high-resolution mass spectrometry methods can be used for characterization of chemical composition of petroleum UVCBs (Palacio Lozano et al. 2020; Rodgers and McKenna 2011). In this regard, by coupling high-resolution mass spectrometry with additional separation techniques, such as GC-APCI (Barrow et al. 2014) or ion mobility (Maillard et al. 2021; Ruger et al. 2021), additional characterization of isomers can be achieved. It is important to distinguish and characterize structural isomers in petroleum UVCBs to understand potential variability in the manufacturing process chemistry and the effects of different oil feed stocks (Lalli et al. 2015).

3.6. Conclusion

This study evaluated samples of 6 petroleum products (heavy aromatic, hydrotreated light paraffinic, and hydrotreated heavy paraffinic) from 2-3 production cycles using GC-MS and APPI(+) IMS-MS. The resulting data were used for classification and grouping using several unsupervised algorithms as either untargeted data, or after structurally related compounds in each sample were identified with confidence using multidimensional data analysis workflow. Between- and within-substance variability was quantified and the types of hydrocarbon blocks, and individual molecules, that were variable in samples of different production cycles were identified. Sample analysis was conducted in different laboratories to examine reproducibility of the grouping and classifications. Overall, these data show that IMS-MS can be used to provide chemical compositional data on petroleum UVCBs, information that is needed to characterize the

variability in substances from different production cycles. Such chemical characterization can be used to support hazard evaluations and address the regulatory need for qualitative and quantitative comparative assessment of the chemical composition of petroleum UVCBs.

3.7. Acknowledgements

This study was supported, in part, by grants from the National Institutes of Health (P30 ES029067 and P42 ES027704) and the National Academies Gulf Research Program (2000008942). A.T. Roman-Hubers was supported, in part, by a training grant from the National Institutes of Health (T32 ES0226568).

Table 3.1 Petroleum refining products used in this study. Samples of the same product (identified by sample ID) are numbered consecutively based on their date of collection. See Table S3.1 for additional information.

Sample ID	CAS #	Name	Substance Definition
AR150 [1] AR150 [2] AR150 [3] AR200 [1] AR200 [2] AR200 [3]	64742-94-5	Solvent naphtha (petroleum), heavy aromatic	A complex combination of hydrocarbons obtained from distillation of aromatic streams. It consists predominantly of aromatic hydrocarbons having carbon numbers predominantly in the range of C_9 through C_{16} and boiling in the range of approximately 165°C to 290°C (330°F to 554°F).
BO60 [1] BO60 [2]	64742-55-8	Distillates (petroleum), hydrotreated light paraffinic	A complex combination of hydrocarbons obtained by treating a petroleum fraction with hydrogen in the presence of a catalyst. It consists of hydrocarbons having carbon numbers predominantly in the range of C_{15} through C_{30} and produces a finished oil with a viscosity of less than 100 SUS at 100°F (19cSt at 40°C). It contains a relatively large proportion of saturated hydrocarbons.
BO100 [1] BO100 [2] BO220 [1] BO220 [2] BO220 [3] BO600 [1] BO600 [2] BO600 [3]	64742-54-7	Distillates (petroleum), hydrotreated heavy paraffinic	A complex combination of hydrocarbons obtained by treating a petroleum fraction with hydrogen in the presence of a catalyst. It consists of hydrocarbons having carbon numbers predominantly in the range of C_{20} through C_{50} and produces a finished oil of at least 100 SUS at 100°F (19cSt at 40°C). It contains a relatively large proportion of saturated hydrocarbons.

Table 3.2 A list of features that that exceeded the thresholds for both abundance of 0.1% and significance (multiple testing-corrected p-value) in three tested products. See Figure 3.8 for additional details.

Product Name	Feature ID*	Relative abundance, % total (mean±SD)	Fold Difference**	-Log$_{10}$ (p-value)***	Inferred formula[#]	Hydro-carbon class	Putative feature identity[†]	REACH indication of concern[¶]
AR150	8	4.3±2.6	3.01	7.89	$C_{12}H_8$	TriAr	Acenaphthylene	Annex III substances
AR200	37	1.6 ± 1.4	10.6	6.28	$C_{15}H_{11}$	TriAr	Methyphenantrene or methylanthracene	Annex III substances
	121	0.80 ± 0.22	26.5	7.66	$C_{17}H_{14}$	NTriAr	Cyclopenteno-phenanthrene	-
	26	0.77 ± 0.39	2.97	6.75	$C_{14}H_{10}$	TriAr	Anthracene	PBT, SVHC
	90	0.66 ± 0.14	1.51	6.22	$C_{12}H_6$	MonoAr	Triethynylbenzene	-
	77	0.35± 0.074	7.94	6.66	$C_{16}H_{18}$[#]	DiAr	Diphenylbutane	-
	341	0.32 ± 0.10	10.6	8.75	$C_{16}H_{16}$[#]	TriAr	Propylfluorene	-
	340	0.28 ± 0.080	5.03	8.53	$C_{15}H_{14}$[#]	TriAr	Ethylfluorene	-
	73	0.24 ± 0.63	11.6	6.78	$C_{17}H_{16}$	TriAr	Trimethylphenanthrene	Annex III substances
	475	0.19 ± 0.049	12	6.34	$C_{17}H_{18}$[#]	NDiAr	Benzyl-tetrahydronaphthalene, or ethyl-methyl-dihydroanthracene	-
	501	0.13 ± 0.033	1.56	6.23	$C_{13}H_{12}$[#]	DiAr	Methyl-phenylbenzene	Annex III substance
BO220	334	0.42 ± 0.095	1.57	6.29	$C_{17}H_{22}$	DiAr	Heptylnaphthalene	-
	32	0.27 ± 0.086	1.65	6.12	$C_{31}H_{50}$	DiAr	Henicosanylnaphthalene	-
	213	0.13 ± 0.0012	12.3	6.27	$C_{14}H_8$[#]	DiAr	Diethynylnaphthalene	-

*, See Supplemental Table 5A for additional information on each feature.

**, The maximum of the absolute value (if at least 3 samples were available) of the fold difference when comparing across all pairs of samples from independent production cycles.

***, The minimum p-value (converted to a -Log$_{10}$) from unequal variance t-test for comparing the differences in abundance of a feature between samples from independent production cycles of a product.

[#], Table 5A lists these features in their radical form (-H)

[†], Putative identification based on the data analysis workflow as detailed in Methods, or based on a match to a library standard (*i.e.*, anthracene).

[¶], The indications of concern for each putatively identified molecule based on the information in ECHA database (https://echa.europa.eu/). PBT, persistent, bioaccumulative, or toxic; SVHC, substance of very high concern; -, no information was included in the database as of 10/2021.

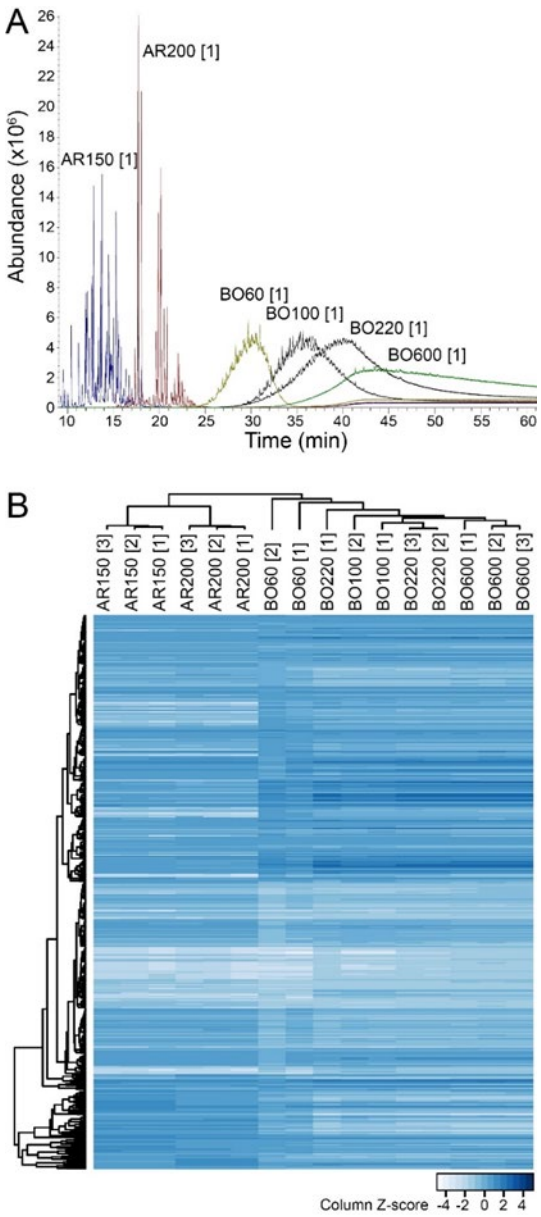


Figure 3.1 GC-MS full scan analysis of petroleum UVCB products included in this study. (A) Superimposed GC-MS total ion chromatograms (time vs. abundance) for representative samples (see Table 3.1 for sample annotations). Individual chromatograms for each sample are shown in Supplemental Figure 1. (B) Hierarchical clustering analysis of the average abundance of the detected compound ion fragments in a mass range of 40-500 amu in 10,127 scans (see Table S3.2 for the raw data).

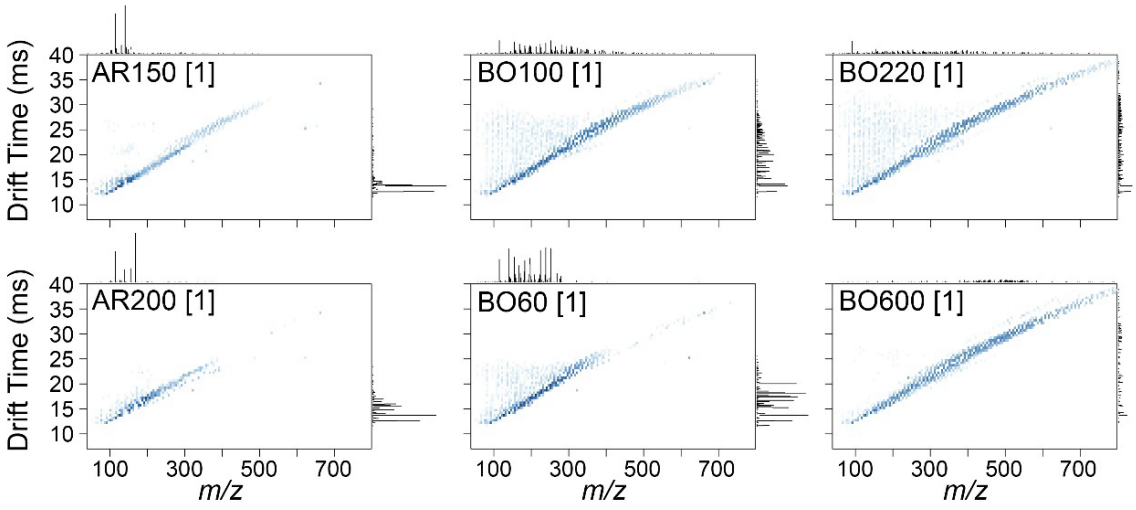


Figure 3.2 Representative nested APPI(+) IMS-MS spectra for petroleum UVCB products included in this study. Representative samples (see Table 1 for sample annotations) are shown, data for other samples are shown in Supplemental Figure 2. Individual features are shown as dots in the 2D scatterplot where x-axes are m/z, y-axes are drift time, and feature intensities are indicated by the color intensity. The density histograms of the features are shown at the top (for *m/z*) or on the right (for drift time) of each plot.

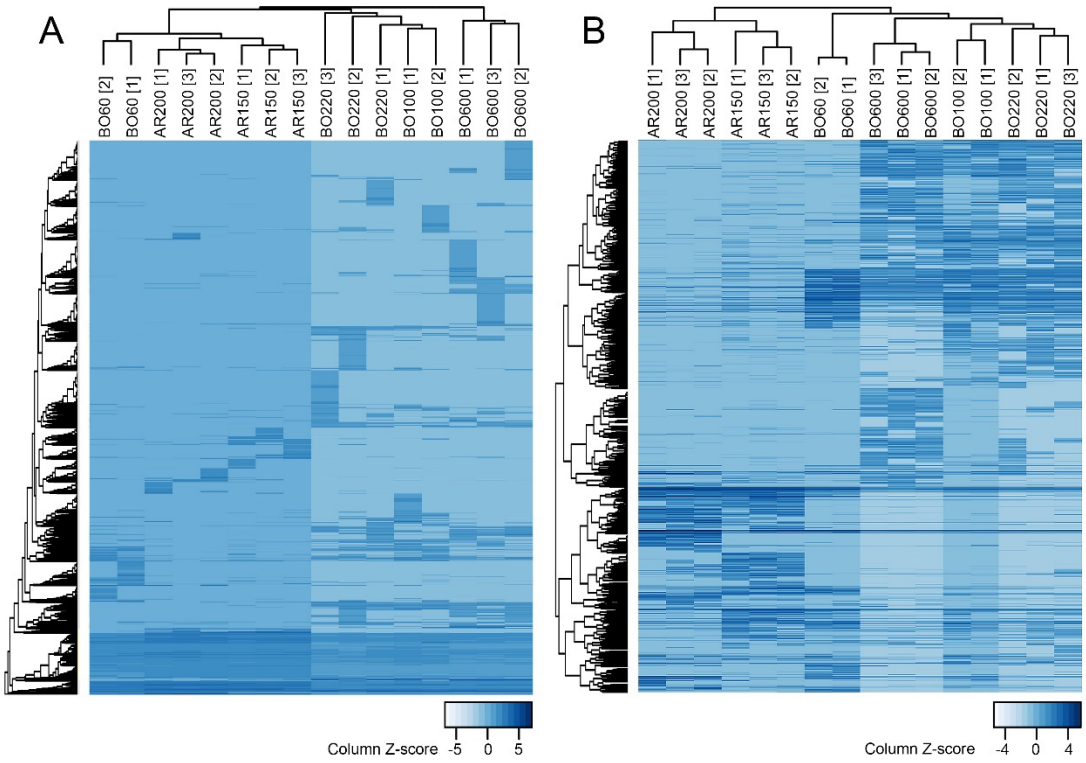


Figure 3.3 Unsupervised hierarchical clustering of petroleum UVCB products using IMS-MS data. Shown are heatmaps (illustrating relative feature abundance) that were products of hierarchical clustering analysis (Spearman correlation, average linkage method) for 16 samples (see Table 3.1 for sample annotations) analyzed in one of the experimental runs. Technical replicates of each sample were averaged for each feature. (A) Full dataset (Table S3.3A; 55,466 features). (B) Filtered dataset (Table S3.4A; 1,530 features).

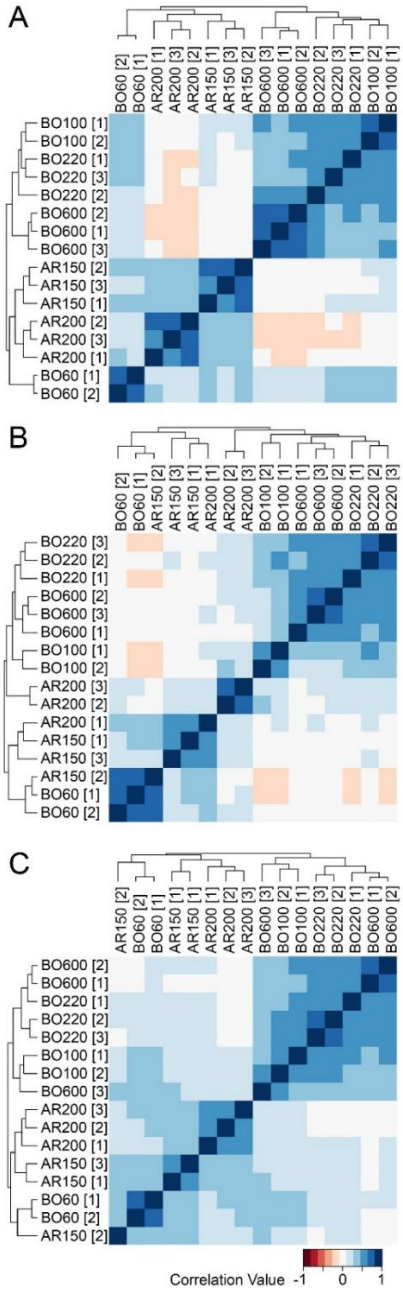


Figure 3.4 Inter- and intra-laboratory reproducibility of grouping petroleum UVCB products using untargeted IMS-MS analyses conducted in independent experiments. The samples were analyzed using an identical experimental protocol either at Texas A&M on the same instrument but by a different operator (A and B) or at North Carolina State University by another operator and instrument, but the same model of IMS-MS platform (C). Correlation values are listed in Table S3.7 and shown using a color gradient as indicated in the legend at the bottom of the figure.

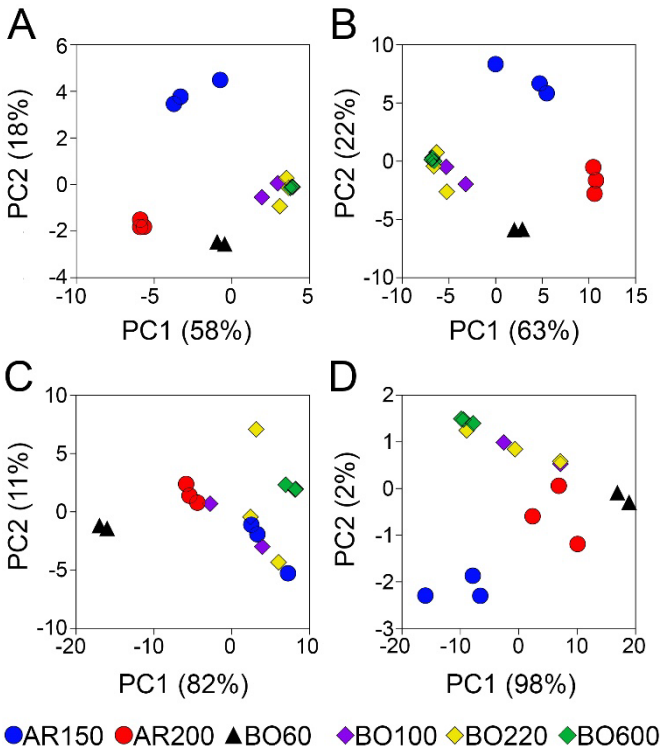


Figure 3.5 The Principal Component Analysis grouping of petroleum UVCB products. (A) Grouping based on the relative abundance of all features with assigned molecular formulas (Table S3.5). (B) Grouping based on the carbon chain length distribution (Table S3.8). (C) Grouping based on the hydrocarbon class (Table S3.8). (D) Grouping based on the heteroatom profile (Supplemental Table 9). Colors represent individual samples of the same product as indicated in the legend at the bottom of the figure.

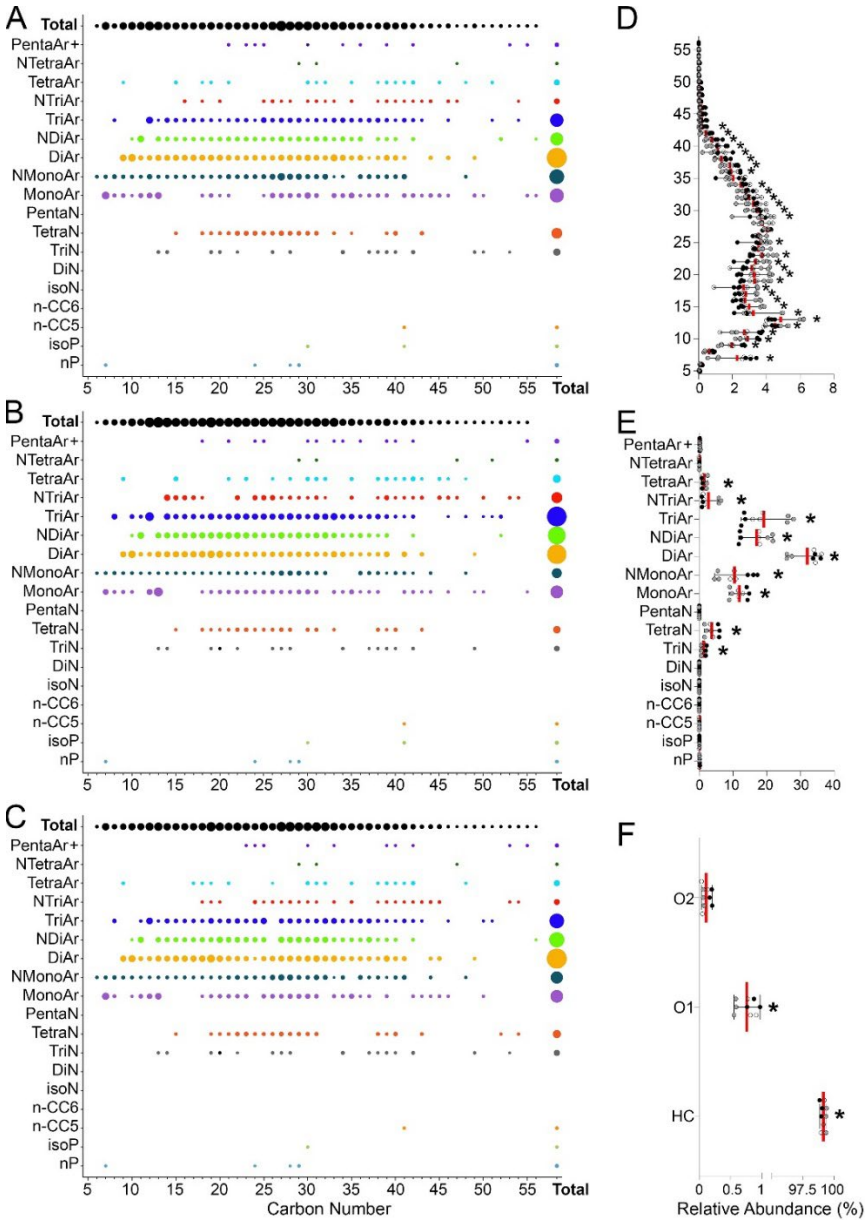


Figure 3.6 Hydrocarbon block matrix for samples from independent manufacturing cycles of product BO220. (A-C) Dot plots representing the relative abundance (each sample is scaled to 100%) of the constituents in different hydrocarbon blocks (hydrocarbon class vs carbon chain length) in three independent samples (see Table S3.8-9 for data on each product). (D-F) Relative abundance distribution for the carbon chain length (D), hydrocarbon class (E) and heteroatom content (F) where symbols represent individual technical replicates (same color) of the samples from independent manufacturing cycles (shades of gray). Red vertical lines are mean and whiskers are min-max range. Asterisks (*) denote blocks with statistically significant (padj-value <0.05, Table S3.10) variability among samples of product BO220 from independent manufacturing cycles.

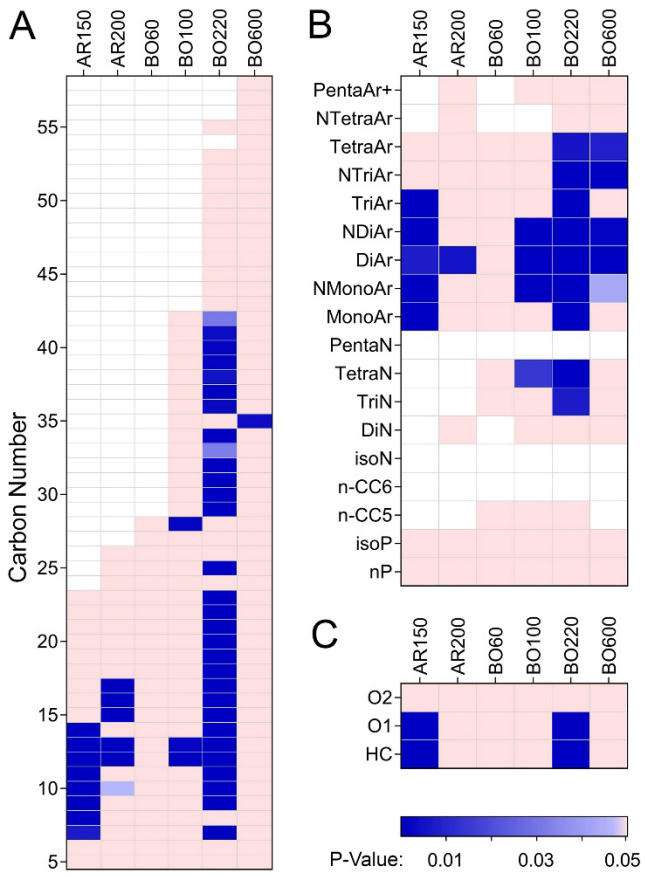


Figure 3.7 Variability in hydrocarbon blocks (A-B) and heteroatom content (C) for independent manufacturing cycles of petroleum UVCB products. Heatmaps show whether relative abundance of the constituents in different hydrocarbon blocks or heteroatom classes were significantly variable (p_{adj}<0.05, see Table S3.9) among samples from independent manufacturing cycles. Colors represent significance (see legend at the bottom of the figure, white indicates that there were no constituents in that hydrocarbon block).

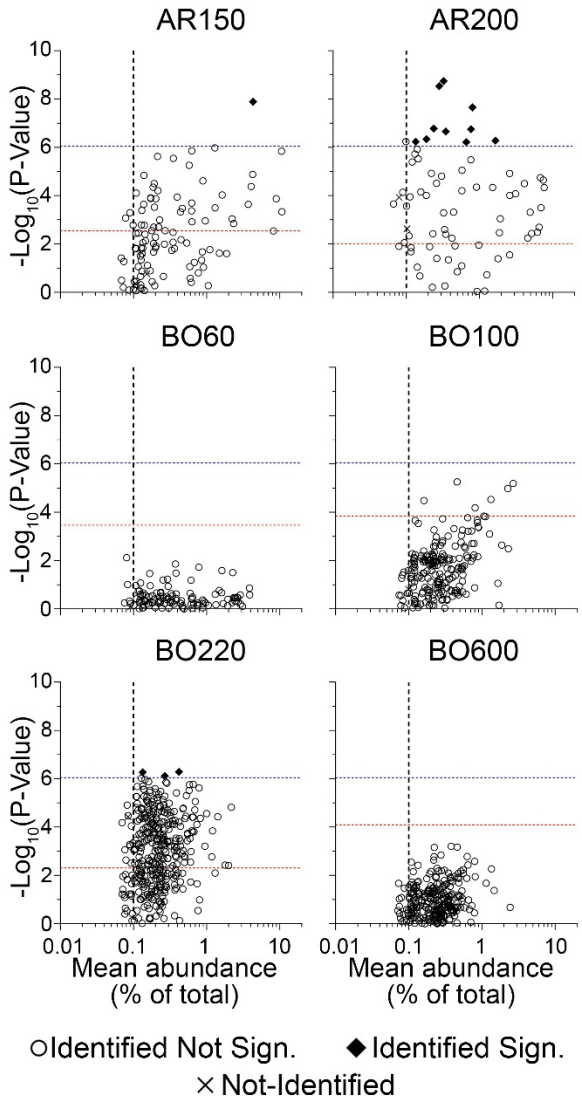


Figure 3.8 Identification of the individual features that are both abundant and significantly variable among samples from independent manufacturing cycles of each petroleum UVCB product. The scatted plots show features that were present in each product based on their relative abundance (x-axis) and significance in variability (y-axis, p-values were converted to $-Log_{10}$(values). Vertical dotted lines indicate the 0.1% relative abundance threshold. Horizontal lines indicate product-specific (red dotted line corresponding to the p-value at false discovery rate of 5%) and global (across all samples, $-Log_{10}$(p-value) = 6.05, blue dotted lines) thresholds for multiple-corrected significance values. Black diamonds indicate features that were exceeding both global variability significance and abundance thresholds (see Table 3.2 for the complete list). Open circles (features with molecular formulae assigned) and "x" symbols (no molecular formulae assigned) indicate features that were not significant based on the global variability significance threshold.

4. DEFINING THE ROLE OF PHOTOOXIDATION AS A WEATHERING PROCESS THROUGH A TIME-COURSE ANALYSIS OF THE WATER-SOLUBLE FRACTION OF AN SUNLIGHT IRRADIATED OIL SLICK

4.1. Overview

Understanding of the contribution of photooxidation to oil weathering is a topic of great interest in both chemistry and regulatory science. Most previous studies evaluated the chemical composition of photooxidized water-soluble fraction of weathered oil under artificial laboratory conditions which presents challenges to translating these findings to real-life oil spill conditions. Therefore, this study used an outdoor mesocosm experiment with crude oil on natural seawater that was constantly circulated under an oil slick (loading of 0.5 g of oil/L of water) and exposed to natural sunlight irradiation. Molecular composition of dissolved organic content from irradiated and non-irradiated samples was quantified over 8 days. Both ion mobility spectrometry mass spectrometry (IMS-MS) and biomimetic extraction-solid phase microextraction followed by gas chromatography-flame ionization detection were used. The results show that oil irradiation led to increased dissolution of oxygenated hydrocarbon fractions in water, primarily O1-containing molecules.

4.2. Introduction

Crude oil may enter aquatic environments as a result of both natural seeps and accidents. Upon exposure to the natural elements, including solar irradiation and microorganisms, chemicals in oil can evaporate, dissolve, disperse, emulsify, biodegrade,

and be photooxidized (National Research Council 2003). Decades of study have been devoted to characterizing oil weathering because these processes influence potential spill impacts and the "window of opportunity" for response decisions (NASEM 2020). Photooxidation is one important weathering process because oil on the water surface is typically exposed to abundant solar radiation; it may involve direct photolysis in which aromatic hydrocarbons absorb light and degrade to photoproducts. Alternatively, a wide variety of hydrocarbons, including non-aromatic components, undergo transformation through reactions with reactive oxygen species produced by solar radiation (Shankar et al. 2015). Recent studies demonstrated that the rate of photooxidation is comparable to other weathering processes (Ward and Overton 2020) and may contribute to an increase in slick oil density that reduces the effectiveness of chemical dispersants (Aeppli et al. 2022; NASEM 2020). In addition, weathering is an important pathway in modulating potential exposures that may result in environmental and human health effects of oil spills because this process can degrade some oil components and generate photoproducts that are more soluble (Wang and Fingas 2003a; Zito et al. 2020). Elucidating how photooxidation alters the composition of oil and its water-soluble fraction is an important research topic since the nature and temporal evolution of this mixture dictates the potential hazards and risks of a spill (Chen et al. 2021; Payne and Phillips 1985; Tarr et al. 2016).

The extent of oil photooxidation depends on the spectrum and irradiance of the light, the light transparency of the surface water, the hydrocarbon composition of the oil, the presence of photo-reactants and scavengers, and the natural movements of the water (National Research Council 2003). Given the complexity of factors that may impact oil

photooxidation in natural environments, studies of chemical reactions and products are difficult. Most published studies of oil photooxidation reproduced sunlight using artificial light sources in the laboratory and lack water movement in their experiments; therefore, there are a number of potential challenges with extrapolating these data to the field (Shankar et al. 2015). In addition, most studies use limited-resolution analytical methods (King et al. 2014; Wang and Fingas 2003a); these analyses provide information on the overall fate of general classes of oil constituents following photoirradiation. Detailed characterization of the water-soluble substances formed from oil during photooxidation has been achieved recently using both high-resolution (Chen et al. 2021; McKenna et al. 2021) and multi-dimensional (Snyder et al. 2021) mass spectrometry approaches.

Indeed, the advancements in analytical petroleomics, such as the use of high-resolution mass spectrometry, have improved understanding of how photooxidation affects the fate and transport of potentially harmful constituents in water (Prince and Walters 2022). For example, Fourier transform ion cyclotron resonance mass spectrometry (FTICR-MS) enables detailed chemical characterization of petroleum substances and has been used to characterize formation of water-soluble polar heteroatom species during photooxidation of oil under laboratory conditions (Benigni et al. 2017; Griffiths et al. 2014; McKenna et al. 2021; Ray et al. 2014; Vaughan et al. 2016; Zito et al. 2020). Improvements in acquisition, processing and analysis of the data from FTICR-MS showed promise for increased throughput of this method (Kooijman et al. 2019), this may be especially critical for analysis of field samples during or after oil spills. Tandem gas chromatography coupled with time of flight mass spectrometry (GC×GC-TOF-MS), a

technique that has good throughput and resolution, has also been used to characterize both non-polar polycyclic aromatic compounds and polar heteroatoms formed upon photooxidation (Lübeck et al. 2020; Snyder et al. 2021). In addition, ion mobility spectrometry coupled to quadrupole-TOF-MS (IMS-MS) is a rapid method for high-throughput analytical characterization of petroleum substances (Fernandez-Lima et al. 2009; Roman-Hubers et al. 2021b; Santos et al. 2015); however, it is yet to be applied to studies of the water-soluble fraction of petroleum substances subjected to photooxidation.

Additional important limitation of previous studies that characterized the complex nature and dynamics of components in the water-soluble fraction of photooxidized oil is lack of anchoring to potential toxicity. Passive sampling methods are commonly used to characterize the bioavailability of various organic contaminants (Gobas et al. 2018). For example, biomimetic extraction using solid phase microextraction fibers coated with polydimethylsiloxane has been used for estimating hazard potential of complex petroleum substances (Hedgpeth et al. 2019; Letinski et al. 2014; Redman et al. 2018b). This method can quantitatively assess the bioavailability of both neutral and acidic compounds in water, an advantage for investigating oil photooxidation because previous studies have demonstrated the formation of both un-ionized (e.g. alcohols, ketones, aldehydes, sulfoxides), as well as carboxylic and sulfonic acid photoproducts (Fathalla and Andersson 2011; Sydnes et al. 1985).

The present work compliments a recent report of temporal changes in slick oil properties and composition using an outdoor mesocosm in the absence and presence of natural sunlight (Aeppli et al. 2022). In this study we aimed to characterize the

composition of the water-soluble photoproducts of oil slick under natural sunlight irradiated. We determined formation of photooxidation products over 8 days using untargeted high-throughput IMS-MS technique with atmospheric pressure photoionization (APPI), followed by identification of structurally-related hydrocarbons and heteroatoms (Roman-Hubers et al. 2021a). Furthermore, the analytical study was augmented with passive sampling analysis using solid phase microextraction biomimetic extraction to gain insights on the ecotoxicity potential of the water-soluble component mixtures that resulted over time.

4.3. Materials and Methods

4.3.1. Mesocosm experiment

The mesocosm experiment was performed at the Ohmsett Oil Spill Testing Facility of the Bureau of Safety and Environmental Enforcement (Leonardo, NJ). The experiment set-up is described in detail elsewhere (Aeppli et al. 2022). Briefly, approximately 1,500 L of seawater (salinity 30 ppt) that was collected in June 2019 from Manasquan Inlet, NJ, USA (40.103166°N, 74.035986°W; South of the Inlet) was continuously re-circulated through a series of six connected polyethylene reservoirs (117 cm long, 51 cm wide, 34 cm deep, grey 65-gal tote BC-4721, Bayhead Products Corp, NH) at a flow rate of approximately 2 L min-1. Each reservoir was divided into four sections (29 cm × 51 cm × 34 cm) with polyethylene dividers that did not fully reach the bottom to prevent wind from pushing all the oil to one side, while allowing water to flow underneath the oil slicks. The reservoirs were connected to a pump and a 330-gal (1.25 m3) high-density polyethylene intermediate bulk container. Inside the contained, temperature was controlled with a

chiller and an air bubbler was used to prevent anoxic conditions. Hoover Offshore Oil Pipeline System crude oil blend (API gravity 35.2) was applied to seawater at a 0.5 g/L loading to form an oil film with a thickness of ~270 µm. The irradiated treatments were placed on the deck of the Ohmsett wave reservoir where they were exposed to natural sunlight ("Irradiated" conditions) at an average daily irradiance of 6.6 kWh m^{-2} day^{-1} (Aeppli et al. 2022) for up to eleven days. During rain events, the irradiated treatments were covered with non-transparent polyethylene lids to avoid disturbance and limit the formation of emulsions. An identical concurrent experimental setup was conducted in an adjacent covered loading dock. These non-irradiated treatments were also covered with non-transparent polyethylene lids which prevented exposure to ambient light ("Non-Irradiated" conditions). Water samples (n=42; 3-4 replicates per time point) were collected from each reservoir through a polyvinylchloride pipe using a glass pipette (to not contaminate water samples with residues of the oil slick) from underneath the floating slick. Water samples were collected at six different time points representing days 0, 1, 4, 5, 6 and 8 after addition of the oil to seawater (Table S4.1). Multiple samples per time point were collected in 20 and 40 mL glass volatile organic analysis-certified amber vials for IMS-MS and biomimetic extraction solid phase microextraction analysis, respectively. Samples for IMS-MS analysis were frozen upon collection and shipped to Texas A&M University. Samples for biomimetic extraction solid phase microextraction were split into two groups for immediate preservation where an equal number were either poisoned (50 ppm $HgCl_2$) or acidified (H_3PO_4 to a pH = 2) then sealed with Teflon® faced septum screw

caps and transported to ExxonMobil Biomedical Sciences Lab in Annandale, NJ where they were refrigerated (4°C) until analysis.

4.3.2. IMS-MS instrumental analysis

Analysis of the water samples was carried out using IMS-MS drift tube-type instrument filled with nitrogen gas (model G6560A; Agilent Technologies, Santa Clara, CA) coupled with an APPI ionization source (model G1917C; Agilent Technologies). The instrument was calibrated prior to sample injections according to the Agilent protocol for 50 – 1,700 *m/z* range using the atmospheric pressure chemical ionization (APCI)-L Low Concentration tuning mix solution (part #G1969-85010, Agilent Technologies). Water samples (1.5 mL) were centrifuged for 5 minutes and the top clear layer (1 mL) was added to 1 mL of methanol (CAS no. 67-56-1, Sigma-Aldrich, St. Louis, MO) mixed with 0.05% of acetic acid (CAS no. 64-19-7, Sigma-Aldrich). Samples were thoroughly vortexed and 200 μL of each sample (n=42) was infused (50 μL/min) directly into the APPI source. Each sample was injected in triplicate and solvent (methanol:acetic acid) washes were conducted between samples to prevent carryover. The ion mobility spectrometry instrumental and ionization source parameters were as follows: APPI positive ion mode, sample analysis time 1.5 min; source parameters: gas temperature 325 °C, vaporizer 350 °C, drying gas 10 L/min, nebulizer 30 psi, VCap 3000, fragment 400 V, 110 RF Vpp 750. The following acquisition parameters were defined for each instrumental run: mass range 50 to 1700 m/z, frame rate 1 frames/s, IM transient rate 18 transients/frame, maximum drift time 60 ms, time-of-flight transient rate 600 transients/IM transient, trap fill time 20,000 μs and trap release time 300 μs. QTOF parameters were as follows: firmware Ver

18.723, Rough Vac 2.71 torr, Quad Vac 3.68 × 10−5 torr, TOF Vac 3.47 × 10−7 torr, drift tube pressure 3.940 torr, trap funnel pressure 3.790 torr, chamber voltage 5.96 μA, and capillary voltage 0.076 μA.

4.3.3. IMS-MS data processing

Raw data files from each experimental run were processed using MassHunter Browser Acquisition data software (ver. 08.00; Agilent Technologies) to derive the nitrogen gas-filled drift tube collisional cross section ($^{DT}CCS_{N2}$) values for all detected features (Stow et al. 2017). The data files for all samples with their respective technical replicates were then uploaded to Agilent MassProfiler software (ver. B.08.00) for feature alignment based on drift time (±5.0%) and mass (±15ppm+5mDa). The aligned raw data matrix (n=6892 features across 42 samples×3 replicates and solvent blank runs) is included as Table S3.2. Then, the raw data for all samples were (i) filtered to remove features with low abundance (i.e., a feature was retained if it had an abundance of >500 in at least two of the three replicates in at least one sample), (ii) the average intensity of each feature present in solvent blank samples was subtracted; and (iii) the average intensity of each feature present in the samples from day 0 was subtracted. If the values resulting from steps (ii) and (iii) above were negative, a small constant (0.001) was entered for that feature. Overall 759 high-abundance features were retained for further analyses after these steps for 34 experimental samples (Days 1-8) and the data from experimental replicates for each of these samples was averaged. These data are available in Table S4.3.

Elemental composition of high-abundance features was further evaluated as follows. First, features were crossed-referenced to a $^{DT}CCS_{N2}$ standard library (Baker

2021); features were assigned a match to a standard in the database with a *m/z* tolerance of ±5 ppm and ±2 mDa and $^{DT}CCS_{N2}$ tolerance of ±1%. Second, we used Kendrick mass defect (KMD) analyses based on CH2 and H functional units to enable compositional analyses as detailed elsewhere (Roman-Hubers et al. 2021a). Overall, putative molecular formulas were assigned using these methods to ~85% (648 out of 759) of the high-abundance features (Table S4.3). Carbon number range and double bond equivalent (DBE) were calculated for each feature that had an assigned putative molecular formula. Feature abundance data were used to evaluate correlations across samples, time and exposure conditions through hierarchical clustering (Euclidean correlation and complete linkage) using the *hclust* package in R studio (Ver. R-4.1.0). Significance of the differences between groups and conditions was evaluated using 2-way ANOVA followed by Sidak's multiple comparison test in GraphPad Prism (ver. 9.0.0; GraphPad Software, San Diego, CA).

4.3.4. Biomimetic extraction analysis

Automated biomimetic extraction-solid phase microextraction analysis was performed on a Perkin Elmer Autosystem gas chromatograph with flame ionization detector. The gas chromatography was equipped Rtx-1 15 m x 0.53 mm x 1.5 µm capillary column with (Catalog# 10167, Restek, Bellefonte, PA) or equivalent and interfaced with a Gerstel MultiPurpose Sampler (CTC Analytics, Zwingen, Switzerland) configured for automated solid phase microextraction injections. The GC inlet was maintained at 280 °C and contained an empty 1 mm id (narrow bore) liner (no glass wool). Automated solid phase microextraction fiber injections were made in the splitless mode with a split time of

3 min. The carrier gas was helium at a constant flow rate of 17 mL/min. The gas chromatography oven was temperature programed from 40 °C for 3 min up to 300 °C at a rate of 45 °C/minute. The flame ionization detector temperature was 300 °C and the detector signal attenuation was set to the -3.

Water samples were automatically extracted with a 1 cm, 30 μm polydimethylsiloxane (0.132 μL polydimethylsiloxane) solid phase microextraction fiber (CAT No. 57309; Sigma-Aldrich, St. Louis, MO) for 100 min at 30 °C with orbital agitation at 250 rpm prior to injection. The fiber was automatically thermally desorbed for 3 min directly in the gas chromatography injection port. The solid phase microextraction fiber was thermally cleaned for at least 60 min at 280 °C and blank temperature-programmed gas chromatography runs were acquired to ensure that a clean chromatographic baseline prior to sample analysis. Calibration was performed by injecting 0.5 μL liquid (solvent) injections using the air-gap technique at three concentration levels, of 2,3-dimethylnaphthalene (20, 100 and 200 mg/mL). The average molar response factor of 2,3-dimethylnaphthalene (CAS no. 581-40-8) was used to convert the measured gas chromatography-flame ionization detection response (total integrated area) to nanomoles of constituents normalized to the volume of PDMS fiber desorbed. Chromatograms were acquired and processed using Perkin Elmer Total Chrom software. Integration parameters were optimized specifically for each sample type to integrate the area under the curve attributable to the SPME extracted sample. This method has a practical quantification limit of 0.5 μmol/mL polydimethylsiloxane. The data from these analyses are included as Table S3.4.

4.4. Results

To closely represent field conditions of a marine oil spill, photooxidation was studied using an open, recirculating outdoor mesocosm. The physical and chemical changes in properties and composition of the oil under these conditions are described in (Aeppli et al. 2022); as previously reported, oxygenation content progressively increased in slick oil during weathering under irradiated conditions. The IMS-MS analyses of the weathered oil showed an increase in ketones and alcohols, consistent with the increase in bulk oxygen content and presence of ketone and carboxyl photo-products that were identified using GC-MS. In this study, IMS-MS analyses were applied to the water samples that were collected over 8 days of this experiments under both irradiated and non-irradiated conditions. Representative IMS-MS nested spectra (Figure 4.1) of these water samples showed presence of numerous chemical species in all analyzed samples. Because natural seawater used in these experiments contains organic molecules before oil was added (Figure 4.1A), these "background" features were subtracted from all experimental samples where weathered oil was present (Days 1 to 8). See Figure S4.1 for the IMS-MS nested spectra for each sample. A large number of high-abundance features were detected in both non-irradiated (Figure 4.1B) and irradiated (Figure 4.1C) samples, even though no organic extraction or sample concentration was performed prior to analysis. This is noteworthy because this study used an open test system at a lower oil loading (0.5 g/L), as compared to other laboratory-based studies of water-soluble fraction under photooxidation conditions that used closed systems with oil loading ranging from ca. 1 to 100 g/L. (Benigni et al. 2017; Griffiths et al. 2014; McKenna et al. 2021; Ray et al. 2014;

Vaughan et al. 2016; Zito et al. 2020). It was observed that the number of molecules detected in irradiated samples, predominantly in the 150-300 *m/z* range, was about twice that in non-irradiated samples (Figure 4.1D), at and beyond day 4.

The patterns in organic molecules detected in water samples were visualized using unsupervised hierarchical clustering (Figure 4.2). High-abundance molecules (n = 759, Table S4.3) observed in at least one of the samples were included in this analysis. The heatmap (higher color intensity indicates greater abundance) shows general concordance in the compositional profiles of the samples that were experimental replicates (i.e., collected at the same time from different parts of the experimental setup). Specifically, replicates of the same condition (day of collection and presence of solar irradiation) clustered together for all groups. Also, three major clusters were evident – samples from all days under non-irradiated condition, those after 1 day of exposure to sunlight, and those that have been exposed to sunlight for 4 to 8 days. Of note is that irradiated samples from day 1 and non-irradiated samples across all days clustered together, indicating higher similarity in their chemical composition. Based on this grouping of samples, several clusters of compounds were identified – those that were highly abundant regardless of solar irradiation (cluster A, 46 features), those that were enriched in non-irradiated samples (cluster D, 9 features), and those that were enriched in irradiated samples (cluster L, 54 features). Molecular composition assignment of these compounds is included in Table S4.3.

Next, the putative molecular formulas of all compounds included in Figure 4.2 were determined using several petroleomic approaches as detailed in the methods. Among

the 759 compounds selected for further analysis, the molecular formula of 648 was assigned with confidence (85.4%, Table S4.3). Overall, most abundant of these were heteroatoms, especially O1- and O2-containing molecules (Figure 4.3A). The fraction of O1-containing molecules was significantly greater in irradiated samples even after 1 day of exposure and further increased by day 8. A similar trend is observed for O2-containinig compounds detected in the irradiated samples significantly increasing from day 1 of exposure to day 8. By contrast, non-irradiated samples significantly lost O2-containing molecules from day 1 to 8 the relative abundance of O2-containing molecules was greatest in non-irradiated samples and showed a decrease from day 1 to day 8. Interestingly, when the compounds that comprised the L cluster (Figure 4.2) were examined (51 of 54 species could be identified), it was observed that O1- and S1- containing compounds significantly increase from day 1 of exposure, this can be correlated back to the innate composition of the evaluated crude oil (Figure 4.3B). Abundance of dissolved hydrocarbons and O2- and S1-containing molecules also significantly increased after 8 days of irradiation time. The temporal trends in these four chemical classes were examined (Figure 4.4). It was found that water soluble fraction for the molecules that are discriminating features for solar irradiation conditions (cluster L in Figure 4.2) is significantly different from non-irradiated conditions after day 1 of exposure. The exception was O1-containing molecules that were significant as early as day 1. The temporal analysis also shows that all of these classes rapidly reached equilibrium because no significant differences were observed among samples collected on days 4 through 8 of exposure to sunlight.

The distribution of double bond-equivalents (DBE) and carbon number ranges was analyzed for hydrocarbons and O1-containing molecules. Figure 4.5 shows the plots for samples collected at days 1 and 8 under either non-irradiated or irradiated conditions. Among hydrocarbon molecules present in water samples (Figure 4.5A), few changes (>18.6% from day 1 non-irradiated) were found over time under non-irradiated conditions, consistent with data shown in Figure S4.2. However, in irradiated samples from day 8, the influence of sunlight was evident as an increase (>113.3% from day 1 non-irradiated) in presence of higher carbon- and DBE-containing molecules. Among O1-containing molecules (Figure 4.5B), most noticeable differences were an increase with time in the fraction of C10- to C20-containing molecules in both conditions; however, in irradiated samples, there was also an increase in higher DBE-containing species. Additionally, it can be observed that the O1-containing compounds in the water-soluble fraction increases around 4.5-fold (>443.2% from day 1 non-irradiated) after irradiation for 8 days. These data collectively confirm previous laboratory-based observations (Ray et al. 2014) that oil photooxidation results in solubilization of higher DBE-containing species.

To compare H/C and O/C ratios for O-containing compounds present in weathered oil and water-soluble fraction thereof, van Krevelen diagrams (Kim et al. 2003; van Krevelen 1950; 1984) were used (Figure 4.6). For weathered slick oil samples at day 11 (Figure 4.6A, data from (Aeppli et al. 2022)), solar irradiation resulted in appearance of both non-aromatic (H/C>1) heteroatoms, as well as generation of progressively more aromatic compounds (H/C<1). Neither slick oil weathered with or without irradiation showed evidence of higher oxygenated molecules consistent with the expected limited oil

solubility of these products and favorable removable via partitioning into the water column, also observed when plotting O/C versus carbon number (Figure S4.2.). By contrast, a higher and wider range O-containing compounds was observed in the water column at day 8 (Figure 4.6B). For the compounds that were present in both irradiated and non-irradiated samples, the majority were alkylated-aromatic ($1.0<H/C<1.5$) molecules. Solar irradiation exposure was associated with an increase in presence of oxygenated compounds (greater O/C ratio), in accord with previous laboratory-based studies of oil photooxidation (Ray et al. 2014).

To provide hazard anchoring to the analytical data, the bioavailability of neutral and negatively charged acid extractable organic acids was evaluated using solid phase microextraction biomimetic extraction. Raw data from these studies are presented in Table S4.4 and summarized in Figure 4.7. Figure 4.7A shows representative chromatograms from biomimetic extraction solid phase microextraction analysis of acidified samples. These plots demonstrate a temporal enrichment of mixture components absorbed to polydimethylsiloxane from the water-soluble fraction of irradiated samples, which appears in the form of a characteristic unresolved complex mixture "hump". Further, the rate of water-soluble components that partition to the fiber over the 8-day test occurred two-fold faster in irradiated samples as compared to non-irradiated acidified samples (Figure 4.7B). By contrast, a similar increase in fiber concentrations that reflects only neutral mixture components was observed in both irradiated and non-irradiated conditions (Figure 4.7C). The lack of difference in biomimetic extraction solid phase microextraction results for non-acidified (i.e., poisoned) samples likely reflects the large

fraction of components that share common features (Figure 4.2) which are present due to oil dissolution and biodegradation processes that are occurring irrespective of light exposure. These results also suggest that the factor of two difference in bioavailability associated with irradiation is attributable to the formation and subsequent partitioning of acidic photoproducts to the fiber. Examination of Figure 4.7B suggests that the contribution of these photoproducts in enhancing fiber concentrations appear to have peaked by day 6. Based on the maximum average fiber concentrations achieved at this time point of 4.9 μmol/mL_{PDMS} (Table S4.4), the mixture components in irradiated water samples would not be expected to cause >50% acute toxicity given the fiber concentrations corresponding to 50% effect, depending on species and endpoint sensitivity, range from 13.6 to 240 μmol/mL_{PDMS} (Redman et al. 2018b).

4.5. Discussion

Response to an oil spill is time-sensitive, therefore understanding the role of weathering process will define the "window of opportunity"(NASEM 2020). Understanding the physical-chemical transformations on the oil composition has been an area of intensive research to elucidate the potential hazards and risks of a spill. Oil spill modeling has limited the role of sunlight irradiation inducing photooxidation, nevertheless during the Deepwater Horizon spill of 2010 photooxidation appeared to play a major role on the physical-chemical composition of the surface oil at a greater rate than expected influencing the oil's environmental fate (Keramea et al. 2021; Ward and Overton 2020). In this present study, we characterized the composition and bioavailability of the dissolved organic constituents from an irradiated oil slick in a mesocosm exposure experiment.

Therefore, defining the role of photooxidation and addressing the correlated exposures characterized through high-resolution IMS-MS and hazards measured with biommetic extraction to define the related risk.

Petroleomic analysis afforded by high-resolution mass spectrometry, when applied to evaluate complex petroleum substances have generated comprehensive molecular compositional information (Palacio Lozano et al. 2020). Compositional transformations of the water-soluble fraction photo-enhanced oil under laboratory conditions have been characterized by FTICR-MS (Benigni et al. 2017; Griffiths et al. 2014; McKenna et al. 2021; Ray et al. 2014; Vaughan et al. 2016; Zito et al. 2020) summarizing the increase in oxygen containing compounds. Rapid and high-resolution characterization afforded by IMS-MS improves analytical efficiency and resolution through multidimensional separation in a millisecond timescale for rapid elemental characterization of the chemical profiles (Grimm et al. 2017; Ibrahim et al. 2015; Ponthus and Riches 2013; Roman-Hubers et al. 2021b). Therefore, this studied aimed to demonstrate high-throughput isobaric and isomeric characterization of the dissolved organic compounds of a photooxidized oil slick. Our choice of APPI (+) IMS-MS measure the temporal transformations of polycyclic aromatic compounds in the water-soluble fraction across eight days of exposure. The comprehensive molecular profiles demonstrated compositional changes after one day of sunlight irradiation. The photo-enhanced dissolution of the oil slick into the water-soluble fraction was characterized through the increase of oxygen and sulfur compounds. The increase of Ox containing species concords with previous laboratory-controlled studies that expose oil slicks to light, highlighting that photooxidation process is dependent on the

oil and thickness of the slick (Keramea et al. 2021). Nevertheless, compositional photo-transformations were observed to stabilized by day 4. By comparing and contrasting, the non-irradiated and irradiated IMS-MS chemical profile, the activity of other weathering processes can be observed to be affecting the physical-chemical composition of the oil slick and therefore, the water-soluble fraction. Based on the observe trends of water-soluble fraction it is derived that photooxidation effects take place rapidly with an increase in water solubility of dissolved organic components and by day four of exposure other physical-chemical transformations are being induced due to exposure to light.

Second, biomimetic extractions measurements serve as surrogate measurements of the total bioavailable hydrocarbons in the water-soluble fraction which can be translated to a defined critical effect concentration for hazard assessment (Redman et al. 2018a). This passive sampling method is based on the premise that partitioning of oil components between water and the site of toxic action in test organisms (i.e. target lipid) is proportional to partitioning between water and polydimethylsiloxane. Thus, when solid phase microextraction fibers are equilibrated with a water sample containing a complex mixtures of oil components, the resulting amount and composition of components absorbed into the fiber serves as a surrogate for internal concentrations in target lipid of biota that dictates toxicity. Further, the total amount of components sorbed to polydimethylsiloxane, regardless of the specific chemical nature and when expressed on molar basis, provides a quantitative measure of the toxic potency of the sample the aqueous concentration, bioavailability and bioconcentration potential of all mixture components are considered. This method has also been extended to quantify and successfully predict the toxicity of

complex mixtures of acid extractable organics (e.g. naphthenic acids) in oil sands process waters (Redman et al. 2018b). This is accomplished by first acidifying water samples prior to SPME analysis in order to protonate these species into the neutral form and promote partitioning to the fiber. While previous work has demonstrated the applicability of passive sampling methods to investigate photoproducts (Forsberg et al. 2014; Llompart et al. 2019) biomimetic extraction has not been applied as an analytical tool to predict the toxicity of photo-irradiated oil. To this effect, this study aimed to demonstrate the relationship between established temporal photooxidation and potential toxicity. The correlation between temporal irradiation and the calculated accumulation in the fiber allowed to establish the progressive bioavailability of dissolved organic acids. This data supports extensive evidence that has shown the photo-enhanced toxicity of oil at low concentrations to aquatic organisms (Barron et al. 2003; Barron et al. 2005; Calfee et al. 1999; Kim et al. 2019; Maki et al. 2001; Pelletier et al. 1997).

This work presents the unique advantages of untargeted IMS-MS analysis in tandem with biomimetic extraction solid phase microextraction gas chromatography-flame ionization detection as major tools for a rapid screening for exposure and risk assessment during the event of an oil spill. The physical-chemical changes such as the dissolution of oil slick therefor change in viscosity, observed in this mesocosm exposure and characterized through IMS-MS will directly affect the selection of response plan and the "window-of-opportunity" for response of an oil spill. Biomimetic extraction solid phase microextraction measurements confirm the composition and the associated

bioavailable hazards in the environment and the contribution of sunlight irradiation and photooxidation.

4.6. Conclusion

Photooxidation has been considered as a secondary weathering process, the events of the Deepwater Horizon oil spill proved the contrary. The present study, evaluates the role of photooxidation as a weathering process and its effect on the environmental fate of an oil spill. Our results are consistent with previous studies evaluating the fate of an oil slick after irradiation. In addition, untargeted IMS-MS provides new evidence of the role of photooxidation on the environmental fate of an oil slick through characterization of the water-soluble fraction. High-resolution data differentiates the elemental composition distribution across eight days of exposure between Non-Irradiated and Irradiated samples. The application of biomimetic extraction solid phase microextraction measurements served as a model for defining the total bioavailable fraction through exposure. The positive trend observed shows the contribution of photooxidation to increase bioavailability, nevertheless the relative changes do not warrant concern of potential aquatic toxicity. By carrying out a mesocosm exposure, this work allowed for the translation of various variables in the process of weathering and highlights the role of photooxidation. Nevertheless, the rate of photooxidation and/or solubilization cannot be defined due to the limited exposure time and scattered sample collection. This work resembled the conditions field conditions during the event of an oil spill even though the process of emulsification of oil was prevented during the exposure. Overall, this study

addresses the practical limitations of an oil spill response through rapid risk assessment of the effect of photooxidation on the realistic oil spill experiment.

4.7. Acknowledgements

This work was supported, in part, by grants from the National Institutes of Health (P42 ES027704) and the National Academies Gulf Research Program (#2000008942). A.T. Roman-Hubers was supported, in part by a training grant from the National Institutes of Health (T32 ES0226568).

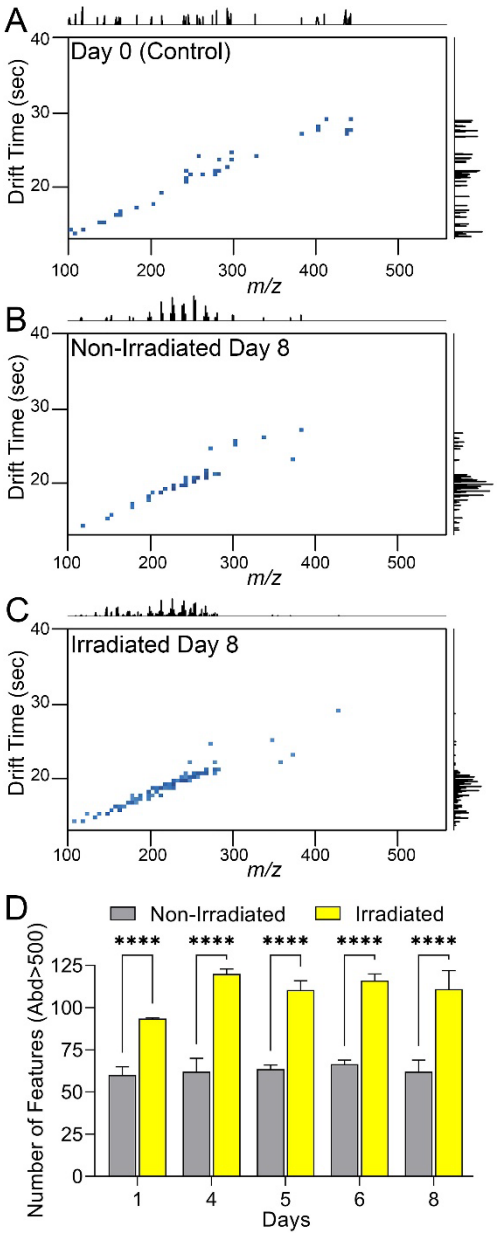


Figure 4.1 APPI (+) IMS-MS spectra (m/z on the x-axis and drift time, parameter used to calculate DTCCSN2, on the y-axis) of representative water samples. Two-dimensional spectra are shown for samples collected on (A) Day 0 (before addition of oil), or (B-C) Day 8 (B, non-irradiated; C, irradiated). Shown are high abundance (see Methods) features that remained after normalization for solvent (in all figures) and water (panels B and C) controls. (D) Total number of unique features with an abundance greater than 500 across all time points (mean and standard deviation for each group as indicated). Asterisks (****) indicate significant difference ($p<0.0001$) between groups.

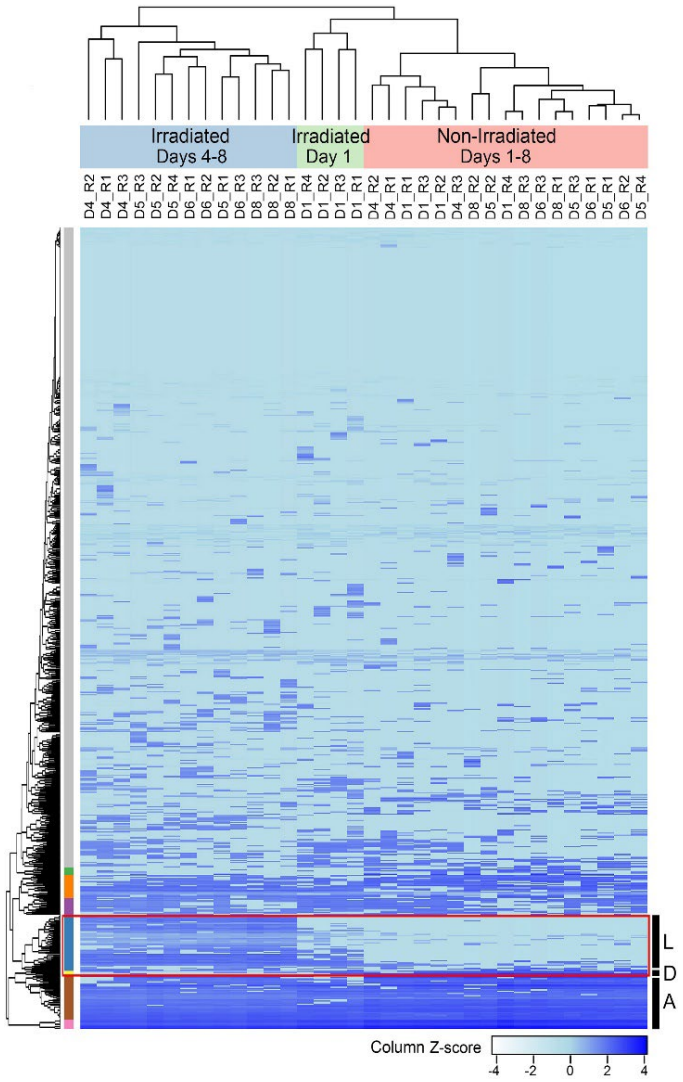


Figure 4.2 Hierarchical clustering analysis of the APPI(+) IMS-MS features. The heatmap shows all high-abundance features (n=759) that were detected in at least one of the samples. Both samples (columns) and features (rows) were subject to hierarchical clustering (Euclidean correlation and complete linkage). Feature abundance was z-scaled (see legend inset) for each sample with lower abundance features indicated by light blue and higher abundance features indicated by dark blue colors. Distinct clusters of samples and features are indicated by color bars. Features that are most discriminating between non-irradiated (D) and irradiated (L) conditions are indicated by the red box. A cluster of most abundant (A) features is also indicated.

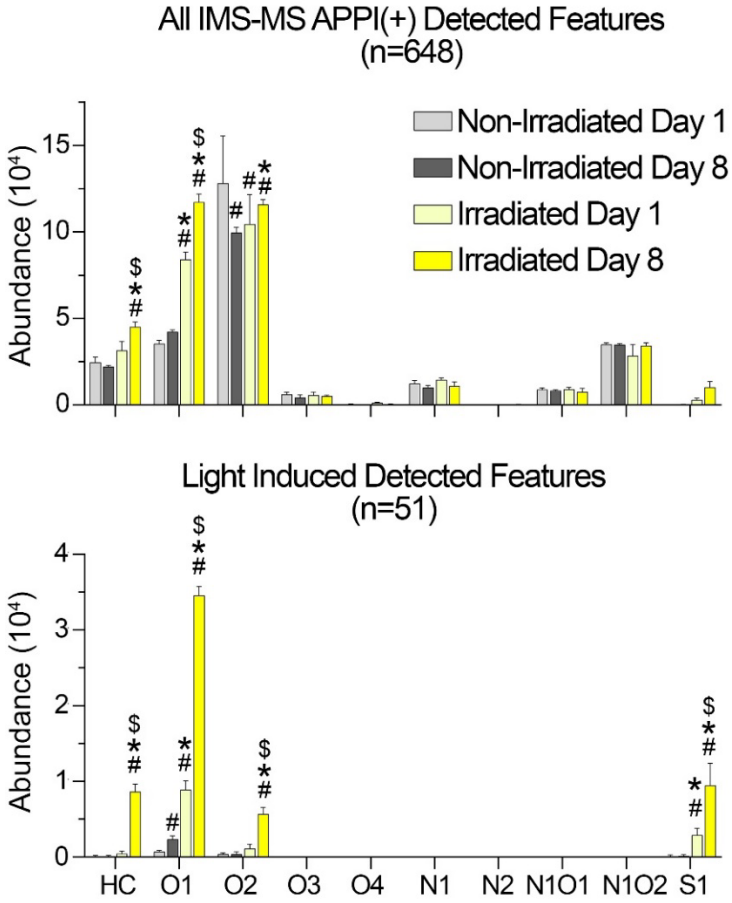


Figure 4.3 Relative abundance in hydrocarbon (HC) and heteroatom class distribution in water samples from days 1 and 8. (A) All high-abundance molecules with putative molecular formulas, or (B) molecules that were present in cluster L (Figure 2), were included in these analyses. Mean and standard deviation is shown for each group (see legend inset for colors). Presence of a symbol above each bar denotes significant (2-way ANOVA with Sidak's multiple comparison analysis, $p_{adj}<0.05$) differences from non-irradiated day 1 (#), non-irradiated day 8 (*), or irradiated day 1 ($).

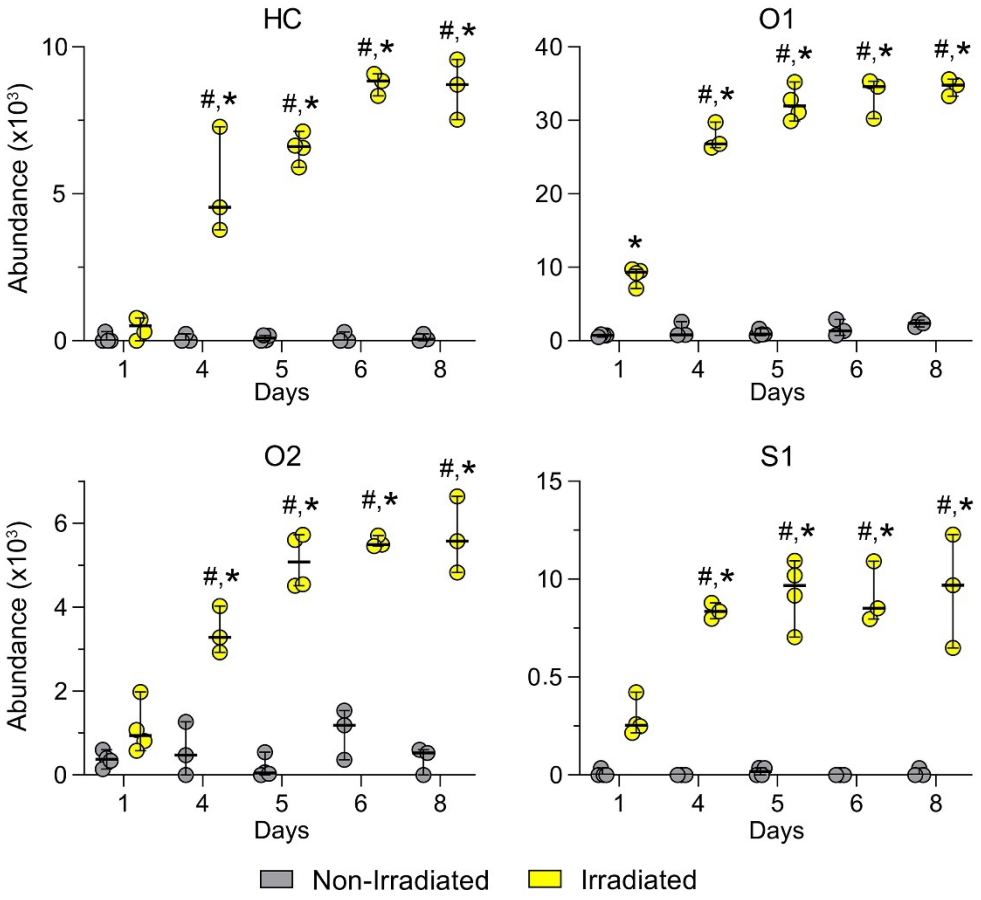


Figure 4.4 Time-dependent changes in relative abundance of hydrocarbon (HC) and heteroatom (O1-, O2- and S1-containing molecules) constituents. Included in these analyses were molecules with putative formula identification (n=51) from cluster L in Figure 2. Feature abundance was averaged for each group (line is mean, error bars are standard deviation and dots are values from the individual samples). Presence of a symbol above each group denotes significant (2-way ANOVA with Sidak's multiple comparison analysis, $p_{adj}<0.05$) differences from irradiated day 1 (#), or non-irradiated group for the same day (*).

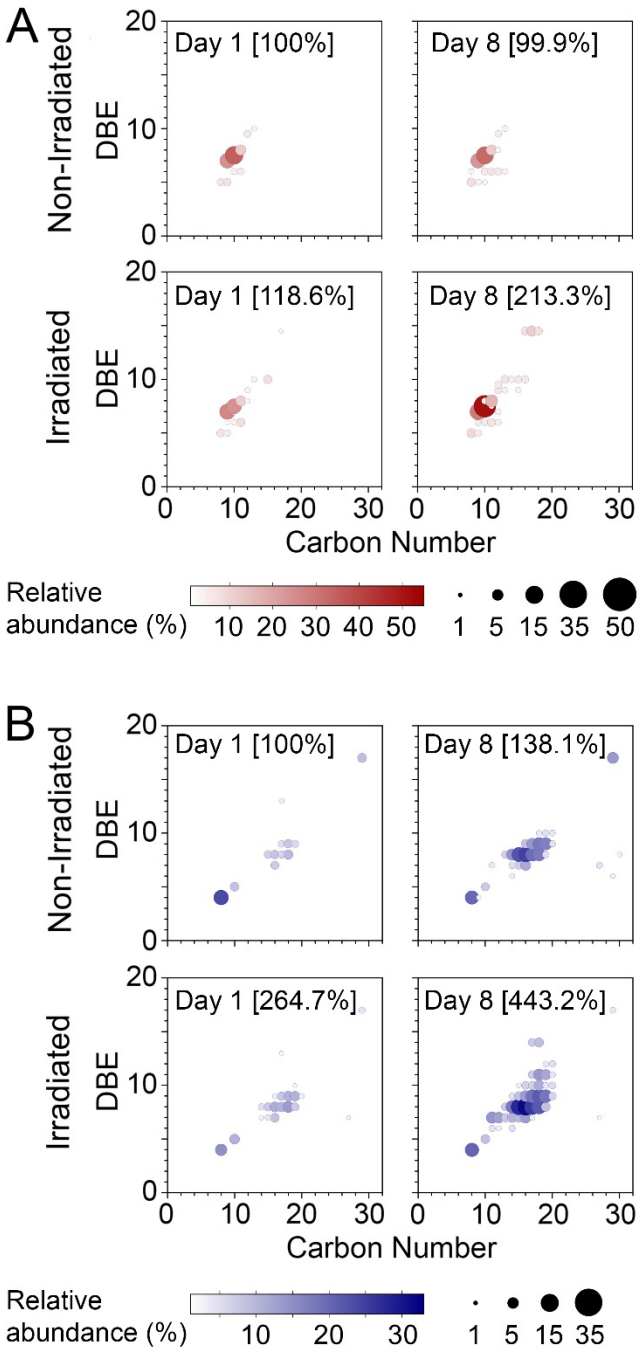


Figure 4.5 Plots of double bond equivalents (DBE) versus carbon number for water-soluble hydrocarbons (A) or O1-containing molecules (B) detected on day 1 or day 8. The relative abundance of the molecules corresponding to each DBE/carbon number in each condition is visualized by the size and color of the circle (see legend) and scaled to the average total abundance of the species in day 1 non-irradiated samples.

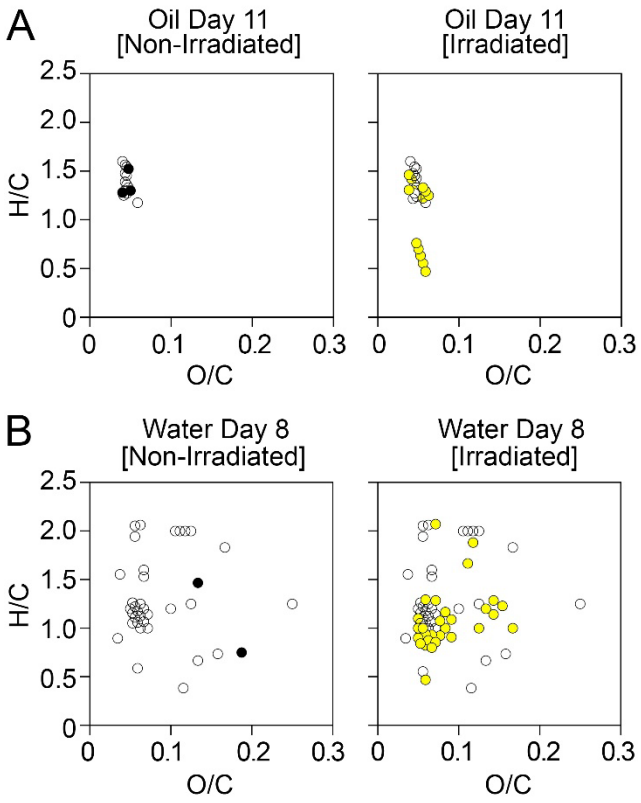


Figure 4.6 Van Krevelen diagrams for molecules generated in (A) oil at day 11, or (B) in water at day 8. Molecules are colored according to their presence in each condition (black circles, present in non-irradiated conditions; yellow circles, present in irradiated conditions; white circled, present in both conditions).

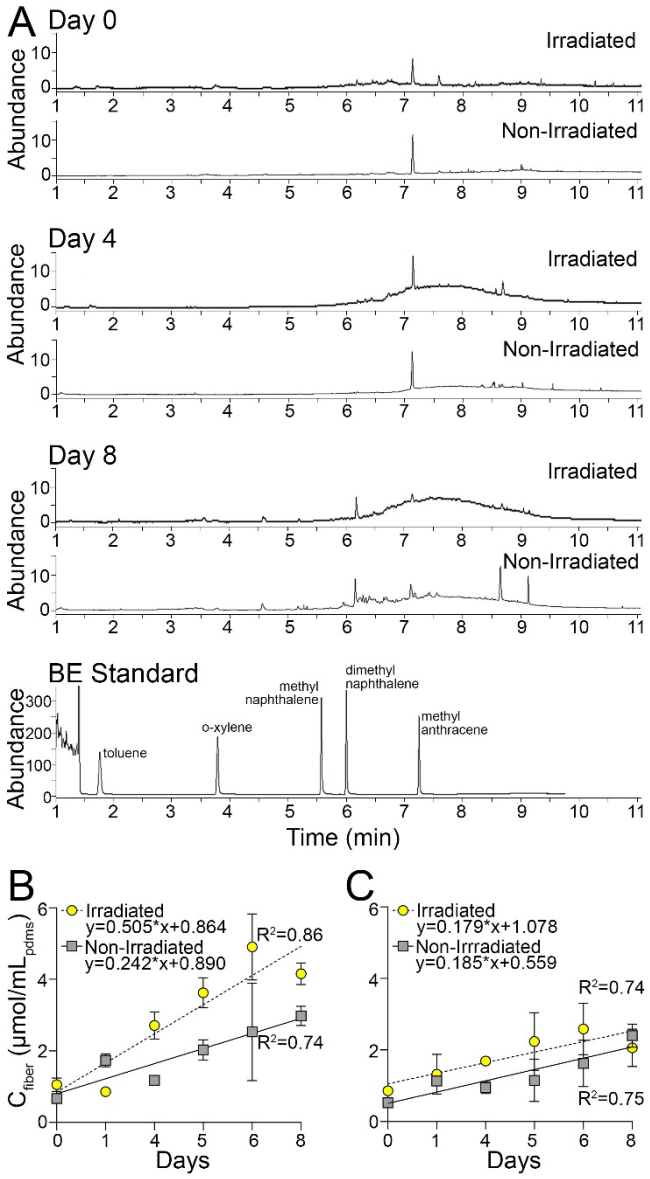


Figure 4.7 BE-SPME analysis of the water samples. (A) Representative GC with flame ionization detection chromatograms for irradiated and non-irradiated samples at various days as indicated. (B-C) SPME fiber concentrations of dissolved organic matter in acidified (B) or poisoned (C) water samples from irradiated (yellow circles) or non-irradiated (gray circles) conditions over the time course of analysis. Shown are mean and standard deviation for each group (see Supplemental Table 4 for numerical values). Linear fit equations and correlations are shown for time-trends in each group.

5. DISCUSSION

5.1. Summary

The inherent compositional complexity and variability of UVCB substances presents challenges when evaluating their hazards, therefore resulting in ambiguous information used to inform substance categories and deduce toxicological properties (Clark et al. 2013; ECHA 2017a). Regulatory agencies and industry partners have raised concern for the need of a deeper compositional understanding of UVCBs for the application of product categorization and read-across in risk assessments (ECHA 2020; European Commission 2021). Advances in analytical high-resolution mass spectrometry technology and petroleomic analysis employed for the characterization of petroleum UVCB substances have efficiently deconvoluted the complex composition at a molecular level resulting in comprehensive molecular information (Headly et al. 2015; Palacio Lozano et al. 2019a; Palacio Lozano et al. 2020). Nevertheless, there is a growing chasm in incorporating comprehensive compositional information afforded by high-resolution mass spectrometry in the assessment of UVCB substances. The work presented in this dissertation aimed to address the prevailing regulatory challenges regarding UVCB substances, by providing an enhanced compositional characterization through novel high-resolution analytics to safeguard appropriate assessment and decision making.

In **Specific Aim 1**, we evaluated the potential application of IMS-MS as a high-throughput technology for the qualitative geochemical characterization and forensic fingerprinting of crude oils. In the event of an oil spill, rapid analytical methodology must

be employed in order to characterize the potential exposures and hazards and identify the source (Stout and Wang 2016; Wang and Stout 2010; Wang and Fingas 2003b; Yang et al. 2015). We directly compared the information generated through GC-MS and IMS-MS for 19 crude oil samples from onshore and offshore areas in the Gulf of Mexico region. In the case of biomarker, gasoline range hydrocarbons and n-alkane measurements obtained through GC-MS analysis each, served to distinguish samples from different regions, but not for samples from neighboring areas. Nearest-neighbor analysis on the chemometric evaluations recognized the additional level of confidence that the rapid, high-resolution IMS-MS data provided. The translational utility of the comprehensive assessment and the value of each data set was evaluated through a science-to-practice tabletop exercise with stakeholders from industry and regulatory agencies. The high-throughput and comprehensive qualitative characterization of IMS-MS was highly praised, likewise the limitations of application of novel method in a real-world scenario were highlighted.

Throughout **Specific Aim 2**, we demonstrate the application of IMS-MS to produce comprehensive molecular profiles of refined petroleum substances addressing the compositional commonalities across products. Regulatory classification of petroleum UVCBs relies on categorization of substances based on physical-chemical commonalities and extrapolation of available information for hazard and risk predictions (Clark et al. 2013; Pusyn et al. 2009). Nevertheless, the broad assumptions of similar chemical composition in the defined categories for read-across practice have been heavily challenged by regulators (ECHA 2020). We demonstrated the ability to assign molecular formulas through the development of a processing workflow combining $^{DT}CCS_{N2}$ and

KMD analysis using CH_2 and H functional units (Roman-Hubers et al. 2021a). By applying the developed and validated molecular characterization workflow, we generated detailed data on the composition of the whole substance and its constituents, as well as data on intrinsic chemical variability of refined petroleum substances. Specifically, in this study we evaluated the compositional information for six petroleum substances obtained from independent production cycles (2-3) from three different manufacturing categories (heavy aromatic, hydrotreated light paraffinic, and hydrotreated heavy paraffinic). Through the identification of structurally related hydrocarbons and heteroatoms, detailed characterization of the most abundant ionizable molecules in a substance was enabled, providing data on between and within-product variability that may inform overall potential hazard. This study demonstrates the application of comprehensive isobaric and isomeric characterization to evaluate broad similarity among substances while also identifying the degree of variability within a class or between production cycles of the same substance. In addition, this study addresses instrumental reproducibility of IMS-MS across operators in the same laboratory and across laboratories, as an analytical method for regulatory decision-making.

In **Specific Aim 3** the environmental fate and behavior of sunlight irradiated oil in a mesocosm experiment was studied through high-resolution characterization of the dissolved organic matter. Millions of gallons of oil enter the ocean and once the oil is exposed to environmental conditions it undergoes several weathering processes inducing physical and chemical changes that define its effects in the environment (Keramea et al. 2021; Nicodem et al. 1997; Payne and Phillips 1985; Tarr et al. 2016). Previous studies

had suggested photooxidation to play a minor role in the weathering of an oil slick; nevertheless, recent data demonstrates it to be an equally active process in the physical and chemical transformations (Nicodem et al. 2001; Payne and Phillips 1985; Ward and Overton 2020). Defining the physical-chemical alterations resulting from photooxidation plays a role in understanding the potential hazards and risk of an oil spill(Chen et al. 2021; Payne and Phillips 1985; Tarr et al. 2016). Through the application of high-resolution IMS-MS we evaluated the physical-chemical changes induced by photooxidation through comprehensive molecular characterization of the oil slick and water-soluble fraction. Furthermore, the ecotoxicity potential of the water-soluble fraction and the impact of photooxidation was identified through biomimetic extraction-solid phase microextraction and measured through gas chromatography-flame ionization detection. In summary, this work showed that sunlight irradiation increases the dissolution of the oil slick resulting in an increase of dissolved organic compounds in the water, primarily, HC and O1 compounds. Together, this suggest that untargeted approaches, such as molecular-level IMS-MS characterization in tandem with biomimetic extraction-solid phase microextraction gas chromatography-flame ionization detection, are significant tools for rapid screening for exposure and risk-assessment during the event of an oil spill.

5.2. Significance

UVCB substances have presented a unique regulatory assessment challenge due to their intrinsic chemical composition (Clark et al. 2013; Salvito et al. 2020). Petroleum substances are prototypical UVCBs, the complex and variable nature of crude oil is further differentiated through manufacturing processes. The inherent UVCB composition of

petroleum substances has been a confounding factor in the assessment of their respective hazards, exposure and risk (Bejarano and Michel 2016; Clark et al. 2013; ECHA 2017a). The limited compositional understanding of petroleum substances has resulted in the reliance on broad physical-chemical and manufacturing properties as commonalities between substances for categorization and toxicological predictions (Grimm et al. 2016). A recent amendment to the to Annex XI, S.1,5 of REACH European Regulation (EC) No 1907/2006 requires that "*Structural similarity for UVCB substances shall be established on the basis of similarities in the structures of the constituents, together with the concentration of these constituents and variability in the concentration of these constituents*" (European Commission 2021). The restricted physical-chemical information defining the "*broadly similar [chemical] composition*" applied for substance categorization and read-across has left a gap in ensuring sound-science risk assessment (ECHA 2020). While much progress has been made in the development of novel analytical technology, there is a growing gap in the application of comprehensive chemical-analytical data in aiding the challenges facing regulatory decision making of UVCB substances. The work presented herein defines a paradigm shift to address the chasm of compositional information regarding UVCB substances through the application of high-resolution analytical data.

5.2.1. Addressing analytical limitations for characterization of UVCB substances through high-resolution mass spectrometry

Chemical analysis of petroleum has been a historically active area of the analytical chemistry and continuous advances have contributed to a better understanding of these

complex substances (Onel et al. 2019; Stout and Wang 2007; Wang et al. 2011). Conventional analytical techniques have been defined through standardized practice for compositional characterization. Nevertheless, these approaches lack the resolving power and sensitivity needed for a comprehensive characterization of UVCB substances (CONCAWE 2012; 2019a). Conventional techniques are limited by resolution capacity and their mass-to-charge range often result in unresolved coeluting compounds; the composition of UVCB substances far exceeds the separation capacity afforded through standardized methods (Wang et al. 2005; Zadro et al. 1985). Additionally, conventional methods rely on extensive sample preparation and large sample quantities prior to lengthy analysis (Stout and Wang 2016). Due to the range in variability of petroleum UVCBs it is not feasible to define a single method for compositional characterization. Therefore, a multi-tier approach has been practiced to facilitate the assessment of these substances. The plethora of assays practiced is also limited to bulk characteristics and volatile organic compounds. Identification of compounds in these complex substances has been limited to prior selected compounds, such as biological markers for oil spill characterization. These limitations have resulted in incomplete understanding of the compositional complexity and variability of UVCBs.

To address these challenges, the advances in high-resolution mass spectrometry have played a critical role in the deconvolution of the complex chemical composition of UVCBs. With the application of IMS-MS resolving power of ~25,000 FWHM and an extensive mass range (>50 *m/z*), we demonstrate across this dissertation the feasibility of acquiring comprehensive chemical profiles for substance categorization and

characterization (Aeppli et al. 2022; Roman-Hubers et al. 2021a; Roman-Hubers et al. 2021b). The separation power afforded by IMS-MS allowed us to resolve closely spaced peaks and heavy molecular weight compounds, which are known to result in unresolved complex mixtures through conventional methods (Fernandez-Lima et al. 2009; Ibrahim et al. 2015; Ponthus and Riches 2013). The post-ionization separation based on drift time enables correlation of the observed mass of an ion to its spatial conformation defined as collisional cross section value (Dodds and Baker 2019). Furthermore, this work collectively demonstrates the general application of high-throughput APPI (+) IMS-MS analysis for the deconvolution of non-polar compounds in crude oil, refined petroleum substances and dissolved organic material (Borsdorf et al. 2006). With little to minimal sample preparation, we demonstrate the application of IMS-MS capacity for qualitative and semi-quantitative characterization of the chemical profile of UVCBs(Grimm et al. 2017). $^{DT}CCS_{N2,}$ ion-specific conformational data allowed us to distinguish isomeric compounds such as branched or linear alkanes and aromatic hydrocarbons (Stow et al. 2017). Incorporation of molecular level characterization, isomeric separation and structural characterization was demonstrated for compositional categorization through hydrocarbon block method (CONCAWE 1996).

5.2.2. Comprehensive time-sensitive characterization for emergency response

Additionally, the significance of high-resolution technology for deducing comprehensive chemical profiles of petroleum of UVCBs was evaluated as an asset for hazard and risk assessment during emergency response. It is well recognized that large quantities of crude oil are released into the environment each year due to accidental spills

or natural seeps; however, both exposure and hazard evaluation of the oil slick and water-soluble fraction remains a formidable challenge to both the industry and the regulators (Laffon et al. 2016). The spilled oil undergoes immediate weathering processes resulting in physical-chemical compositional transformations which will define its environmental fate and effects (Keramea et al. 2021; Nicodem et al. 1997; Nicodem et al. 2001; Payne and Phillips 1985; Tarr et al. 2016; Ward and Overton 2020). Standardized geochemical profiling of crude oil has been assessed through measurement of biological markers to define the compositional environment (Daling et al. 2002; El-Sabagh et al. 2018; Radovic et al. 2012). Therefore, through the work presented herein we present the advancements in petroleomics and chemometrics through IMS-MS for the comprehensive characterization of petroleum substances to aid forensic fingerprinting and fill in the chasm in understanding how weathering affects the fate and transport of oil. Collectively, this work defines the utility of high-throughput, untargeted qualitative analysis for forensic fingerprinting and geochemical profiling of petroleum substances. Narrowing down to the context of compositional differences, chemometric approaches can translate molecular level information to define compositional trends and relationships. This work demonstrates the feasibility to define depositional environments and compositional changes through unsupervised clustering of the comprehensive IMS-MS chemical profile. Moreover, the rapid and high-resolution capacity of IMS-MS for molecular characterization provides detailed information when developing action plans for emergency response.

5.2.3. Improving regulatory assessments through comprehensive compositional information

Likewise, the use of high-resolution IMS-MS for comprehensive chemical profiling of UVCBs can aid regulatory decision making. Regulatory registration of UVCBs has relied on pragmatic approach for assessment of the potential risk and hazards through the categorization of petroleum substances defined by their compositional similarities correlating to similar physiochemical and toxicological properties (Clark et al. 2013; McKee et al. 2015; OECD 2014). The broad physical-chemical properties used for substance categorization were recently challenged by the European Chemical Agency in the application of read-across for new substance registration (ECHA 2020). Throughout this dissertation we demonstrate the value of multidimensional IMS-MS data as a solution to current challenges in chemical characterization of UVCBs for regulatory decision-making purpose. This work demonstrated the feasibility of providing qualitative and quantitative information on the composition of representative complex substances. As well, it shows the translation of comprehensive molecular characterization of the composition of UVCBs to measure variability across substances through chemometric approaches. By summarizing the comprehensive molecular data based on structurally related compounds, the multidimensional chemical profiles enable data translation for sound science decision making (Bierkens and Geerts 2014; CONCAWE 1996; 2019b).

Collectively, this dissertation defines a pragmatic approach for advancing regulatory decision making of UVCB substances through high-resolution analytical chemistry. Comprehensive substance characterization and identification is a critical first

step in validating commonalities between substances for the purpose of read-across predictions. Through the work presented, we demonstrate high-resolution IMS-MS enabled analysis provided molecular level composition characterization of petroleum UVCBs used to establish sufficient information on the composition, as well as to measure trends in compositional changes. Through the integration of novel analytical techniques, we aimed to address the ongoing challenges in the compositional and toxicological evaluation of UVCBs, aid regulatory decision making and safeguard the quality of human health risk assessment.

5.3. Limitations

The studies compiled in this dissertation have successfully evaluated the complex chemical composition of petroleum UVCB substances through the application of IMS-MS. Implementation of this technique and approach for emergency response and risk assessment requires the optimization of current experimental limitations.

In Chapter II (**Specific Aim 1**), we carried out untargeted chemical characterization of 19 crude oil samples from the Gulf of Mexico region. Considering the worldwide upstream industry, the samples screened in this study were limited by source rock and geographical genesis. This study evaluated the chemical profiles for forensic fingerprinting through qualitative chemometric analysis, but lacked a comprehensive feature identification and quantitative analysis. To address these challenges, a library of chemical standards (i.e. biomarkers) with their corresponding $^{DT}CCS_{N2}$ would enrich current feature characterization and identification. Semi-quantification analysis of

constituents in a sample can narrow the gap between current qualitative analysis and desired quantification.

In Chapter III (**Specific Aim 2**), comprehensive molecular characterization of petroleum refined products aimed to measure the degree of commonality across production and substance categories. Due to the complex nature of petroleum substance and the limited range of APPI (+) for ionization non-polar compounds, this study was based on the aromatic (i.e. PAHs) composition. The analysis conducted did not, for the most part, yield data on constituents of low molecular weight or polar composition. Future experimental designs might address the aliphatic and/or heteroatomic sample fractions through the application of other sample ionization techniques (i.e. APCI and ESI) (Zheng et al. 2018). Likewise, additional evaluation on low abundance constituents could provide a deeper insight on overall composition considering the work presented here focused on high quality features based on the REACH abundance threshold of 0.1% (w/w) for notification of constituents classified as "very high concern (ECHA 2017b) .

In Chapter IV (**Specific Aim 3**), the effect of sunlight on the fate of a spilled oiled was assessed through a mesocosm experiment and characterization of the water-soluble fraction through high-resolution IMS-MS. Photooxidation of surface oil and its partitioning to the water column as a function of exposure of time was studied for a period of eight days. As photooxidation has been a slow weathering process, the limited exposure time-frame might have missed ongoing reactions (Nicodem et al. 2001; Payne and Phillips 1985; Ward and Overton 2020). Additionally, sample collection was irregular, not allowing one to define a trend of compositional changes. This study did not include a day-

to-day comparison of the oil slick versus water soluble fraction composition, due to limited samples. In contrast to real-world scenarios, this study was performed under calm conditions, without wave or energy input reducing the potential of emulsification of the oil. Future mesocosm experimental designs could address these gaps by extending experimental period, daily sample collection of water and oil slick and observation of other weathering patterns.

5.4. Future Directions

The intrinsic complex and variable composition of crude oils and petroleum substances is a prominent challenge faced by industry and regulators in decision making. The studies in this dissertation, as well as previously published literature have applied petroleomics to successfully access the previously limited characterization of petroleum substances at a molecular level. Our studies have provided an insight on the capacity of IMS-MS and future applications of these technologies to answer the remaining challenges presented by UVCB substances.

5.4.1. Automation of data analysis workflow to develop a user-friendly software application

IMS-MS data offers an untargeted and multidimensional analysis that when combined with Kendrick Mass Defect (KMD) enables compositional characterization of UVCB substances. Throughout the studies addressed in this dissertation, chemometric and statistical analysis have been defined to summarize IMS-MS data. Our second study presents a data processing workflow that can be applied to characterize the molecular composition through application of a standard reference library, KMD and $^{DT}CCS_{N2}$. In

contrast of IMS-MS rapid assay capacity, these data visualization tools have proven to be time consuming.

The automation of the proposed data analysis workflow would facilitate the high-throughput IMS-MS deliverables. Subsequent use of the scripts created through R Studio (ver. R-4.1.0) using the Shiny package would enable multiuser analysis of IMS-MS data through an interactive app. An automated platform can be utilized to provide elemental composition assignment through (KMD) calculations, data visualization (i.e. hydrocarbon block plots), chemometric analysis (i.e. heatmaps) and statistical analysis (i.e. principal component analysis). This contribution would give access to a diverse background of users to rapid decision-making data on the complex composition of UVCB substances.

5.4.2. Comprehensive chemical profile library of petroleum substances for source identification

Through our first specific aim we presented IMS-MS methodology for grouping complex petroleum substances based on their geochemical profile. Nevertheless, due to the millions of barrels of oils that enter the environment due to natural seeps or accidents, the application of this methodology is limited to the samples we compared with. To employ high-throughput categorization there is a need of a comprehensive library of petroleum substances for direct comparison.

The comprehensive molecular information afforded by high-resolution mass spectrometry allows one to measure the degree of variability across samples and correlate substances by their chemical profiles. Therefore, through a global library encompassing >300 crude oils and refined petroleum products, there is the ability to confidently

characterize and categorize unknown petroleum substances. Molecular characterization through petroleomic analysis can facilitate compositional information for qualitative and quantitative screening of compositional distributions. The comprehensive insights combined with chemometric analyses will allow one to define compositional trends and relationships across samples for thorough comparison for source identification

5.4.3. Application of IMS-MS to source identification of weathered crude oil

The first study of this dissertation carried out forensic fingerprinting of crude oils, to demonstrate its high-throughput capacity for chemical characterization. However, once oil is released to the environment it goes through several transformation processes known as weathering. As a result, the ability of IMS-MS to confidently categorize fresh crude oil and the weathered oil from a spill has to be evaluated.

Evaluation of the chemical composition of fresh versus weathered crude oil through IMS-MS will address this limitation. A direct solution is the ability to generate high-resolution diagnostic compositional patterns of the weathered sample for comparison to the crude oil. Furthermore, the application of chemometric analysis of the acquired profiles can be used for qualitative nontargeted grouping and identification of the source oil. Alternatively, the characterization and semi-quantification of biological markers through their corresponding $^{DT}CCS_{N2}$ can provide information on the oil's depositional environment, organic matter origin, maturity and degree of weathering (Wang and Stout 2010). Measurements of the degree of weathering can also lead to more accurate forensic fingerprinting.

5.4.4. High-resolution IMS-MS data and high-content in vitro read-across for categorization of bioactivity

Hazard characterization of UVCB substances is predicted based on its physical-chemical correlation to defined substance categories. A representative substance known as "worst case scenario" is used to extrapolate hazard data for substances in a defined category. The detailed chemical composition data achieved through high-resolution mass spectrometry facilitate a novel understanding on the potential drivers of hazards in a substance and can define the degree of compositional similarity in a substance category. Nevertheless, the hazard characterization still relies on extrapolated data from *in vivo* studies when no data are available for a specific substance.

High-resolution mass spectrometry data can be coupled with high-content in-vitro data for rapid hazard evaluation of UVCB substances. Such approaches have been demonstrated using a compendium of induced pluripotent stem cell derived in-vitro models representing various tissues of interest (i.e. hepatocytes, cardiomyocytes) for a phenotypic screening of dimethyl sulfoxide soluble extracts (Chen et al. 2020; Grimm et al. 2016; House et al. 2021). ToxPi integration offers data translation of the bioactivity trends observed for quantitative clustering of petroleum substances (Marvel et al. 2018). Therefore, combining bioactivity parameters across multiple cell types with high-resolution mass spectrometry data can reduce data gaps and improve grouping of petroleum substances for hazard evaluation and regulatory approval.

5.4.5. Alternative dosing methods for in-vitro assessment of UVCB substances

While much progress has been made in the development of alternative testing methods to traditional in vivo assays, the hydrophobic characteristics of petroleum substances among its inherent complexity and variability has presented a considerable challenge for standardized *in vitro* methods (i.e. IP 346). Petroleum substances have been subjected to dimethyl sulfide extraction to concentrate the biological fraction (House et al. 2021; Luo et al. 2020b). However, Luo et al showed that dimethyl sulfoxide selectively extracts the polycyclic aromatic compounds, limiting the compositional profile for bioactivity testing (Luo et al. 2020b).

The Organization of Economic Cooperation and Development has provided guidance on "difficult-to-test substances" for aquatic ecotoxicity assays with standardized alternatives for sample dosing (OECD 2019). The methods readily applied from a large-scale ecotoxicity testing, such as water accommodated fraction and passive dosing, can be modified and translated in to a small-scale *in vitro* assay. Therefore, through the application of alternative dosing methods for cell-base assessment, it may be feasible to determine the final bioavailable bioactive constituents of UVCBs.

5.5. Conclusion

In summary, the work presented in this dissertation established IMS-MS as a useful novel analytical tool that can be used for the comprehensive molecular characterization of petroleum UVCB substances. We showed that IMS-MS provides information that can be used not only to evaluate broad similarities among complex samples, but also to determine the molecular composition with high confidence. Together,

these data address the various needs of decision-makers, such as the need for rapid fingerprinting and the need for characterizing the variability in composition. By exploring the advantages and limitations of IMS-MS as an approach for chemical characterization of petroleum UVCBs, we present a sensible strategy that may be implemented within a tiered risk assessment framework designed to ensure sufficient protection from potential human health and environmental hazards. We show that this technique, while not the most high-resolution mass spectrometry technology, can provide sufficient comprehensive molecular compositional information that can identify hazardous constituents, define molecular level commonality, and determine relevant physical-chemical properties of complex substances. Collectively, these data will increase confidence in read-across and the use of other new approach methodologies (i.e. *in vitro* bioactivity data) for risk assessment purposes.

References

Aeppli C, Mitchell DA, Keyes P, Beirne EC, McFarlin KM, Roman-Hubers AT, Rusyn I, Prince RC, Zhao L, Parkerton TF et al. 2022. Oil irradiation experiments document changes in oil properties, molecular composition, and dispersant effectiveness associated with oil photooxidation. submitted.

Aeppli C, Reddy CM, Nelson RK, Kellermann MY, Valentine DL. 2013. Recurrent oil sheens at the deepwater horizon disaster site fingerprinted with synthetic hydrocarbon drilling fluids. Environmental science & technology. 47(15):8211-8219.

Alam MS, Zeraati-Rezaei S, Liang Z, Stark C, Xu H, MacKenzie AR, Harrison RM. 2018. Mapping and quantifying isomer sets of hydrocarbons (≥ c₁₂) in diesel exhaust, lubricating oil and diesel fuel samples using gc × gc-tof-ms. Atmospheric Measurement Techniques. 11(5):3047-3058.

API. 1983. Petroleum process stream terms included in the chemical substances inventory under the toxic substances control act (tsca). In: Control HSRcTFoTS, editor.: American Petroleum Institute.

ASTM International. 2011. Uop method 990-11: Organic analysis of distillate by comprehensive two-dimensional gas chromatography with flame ionization detection. West Conshohocken, PA: ASTM INternational.

Ayala-Cabrera JF, Galceran MT, Moyano E, Santos FJ. 2021. Chloride-attachment atmospheric pressure photoionisation for the determination of short-chain chlorinated paraffins by gas chromatography-high-resolution mass spectrometry. Anal Chim Acta. 1172:338673.

Collision cross section database. 2021. [accessed December 15, 2020]. https://brcwebportal.cos.ncsu.edu/baker/.

Ball GI, Aluwihare Ll. 2014. Cuo-oxidized dissolved organic matter (dom) investigated with comprehensive two dimensional gas chromatography-time of flight-mass spectrometry (gc × gc-tof-ms). Org Chem. 75:87-98.

Barron MG, Carls MG, Short JW, Rice SD. 2003. Photoenhanced toxicity of aqueous phase and chemically dispersed weathered alaska north slope crude oil to pacific herring eggs and larvae. Environ Toxicol Chem. 22(3):650-660.

Barron MG, Carls MG, Short JW, Rice SD, Heintz RA, Rau M, Di Giulio R. 2005. Assessment of the phototoxicity of weathered alaska north slope crude oil to juvenile pink salmon. Chemosphere. 60(1):105-110.

Barrow MP, Peru KM, Headley JV. 2014. An added dimension: Gc atmospheric pressure chemical ionization fticr ms and the athabasca oil sands. Anal Chem. 86(16):8281-8288.

Bejarano AC, Michel J. 2016. Oil spills and their impacts on sand beach invertebrate communities: A literature review. Environ Pollut. 218:709-722.

Benassi M, Berisha A, Romao W, Babayev E, Rompp A, Spengler B. 2013. Petroleum crude oil analysis using low-temperature plasma mass spectrometry. Rapid Commun Mass Spectrom. 27.
Benigni P, Sandoval K, Thompson CJ, Ridgeway ME, Park MA, Gardinali P, Fernandez-Lima F. 2017. Analysis of photoirradiated water accommodated fractions of crude oils using tandem tims and ft-icr ms. Environmental science & technology. 51(11):5978-5988.
Bierkens J, Geerts L. 2014. Environmental hazard and risk characterisation of petroleum substances: A guided "walking tour" of petroleum hydrocarbons. Environ Int. 66:182-193.
Borsdorf H, Nazarov EG, Miller RA. 2006. Atmospheric-pressure ionization studies and field dependence of ion mobilities of isomeric hydrocarbons using a miniature differential mobility spectrometer. Anal Chim Acta. 575(1):76-88.
Calfee RD, Little EE, Cleveland L, Barron MG. 1999. Photoenhanced toxicity of a weathered oil on ceriodaphnia dubia reproduction. Environ Sci Pollut Res Int. 6(4):207-212.
Castiblanco JEB, Carregosa JC, Santos JM, Wisniewski A. 2020. Molecular behavior assessment on initial stages of oil spill in terrestrial environments. Environ Sci Pollut Res. 28:13595-13604.
Chen H, McKenna AM, Niles SF, Frye JW, Glattke TJ, Rodgers RP. 2021. Time-dependent molecular progression and acute toxicity of oil-soluble, interfacially-active, and water-soluble species reveals their rapid formation in the photodegradation of macondo well oil. Sci Total Environ.151884.
Chen Z, Liu Y, Wright FA, Chiu WA, Rusyn I. 2020. Rapid hazard characterization of environmental chemicals using a compendium of human cell lines from different organs. ALTEX. 37(4):623-638.
Cho Y, Ahmed A, Islam A, Kim S. 2015. Developments in ft-icr ms instrumentation, ionization techniques, and data interpretation methods for petroleomics. Mass Spectrom Rev. 34(2):248-263.
Cho Y, Witt M, Jin JM, Kim YH, Nho NS, Kim S. 2014. Evaluation of laser desorption ionization coupled to fourier transform ion cyclotron resonance mass spectrometry to study metalloporphyrin complexes. Energy & Fuels. 28(11):6699-6706.
Clark CR, McKee RH, Freeman JJ, Swick D, Mahagaokar S, Pigram G, Roberts LG, Smulders CJ, Beatty PW. 2013. A ghs-consistent approach to health hazard classification of petroleum substances, a class of uvcb substances. Regulatory toxicology and pharmacology : RTP. 67(3):409-420.
CONCAWE. 1996. Environmental risk assessment of petroleum substances: The hydrocarbon block method. Brussels, Belgium.
CONCAWE. 2012. Reach – analytical characterisation of petroleum uvcb substances. Brussels, Belgium.
CONCAWE. 2014. Guidance on reporting analytical information for petroleum substances in reach registration dossiers v2. Brussels.

CONCAWE. 2019a. Concawe substance identification group analytical program report (abridged version). Brussels, Belgium.
CONCAWE. 2019b. Investigating the hcbm – gcxgc relationship: An elution model to interpret gcxgc retention times of petroleum substances. Brussels, Belgium.
CONCAWE. 2019c. Reach roadmap for petroleum substances. Brussels, Belgium.
CONCAWE. 2020a. Guidance to registrants on methods for characterisation of petroleum uvcb substances for reach registration purposes. Brussels, Belgium.
CONCAWE. 2020b. Hazard classification and labelling of petroleum substances in the european economic area – 2020. Brussles, Belgium.
Petroleum substances and reach. 2021. [accessed]. https://www.concawe.eu/reach/petroleum-substances-and-reach/.
Reach background. 2022. [accessed]. https://www.concawe.eu/reachpost/background/.
Corilo YE, Podgorski DC, McKenna AM, Lemkau KL, Reddy CM, Marshall AG, Rodgers RP. 2013. Oil spill source identification by principal component analysis of electrospray ionization fourier transform ion cyclotron resonance mass spectra. Anal Chem. 85(19):9064-9069.
Covas TR, Santos de Freitas C, Tose LV, Valencia-Davila JA, Rocha YDS, Rangel MD, Cabral da Silva R, Vaz BG. 2020. Fractionation of polar compounds from crude oils by hetero-medium pressure liquid chromatography (h-mplc) and molecular characterization by ultrahigh-resolution mass spectrometry. Fuel. 267.
D'Andrilli J, Dittmar T, Koch BP, Purcell JM, Marshall AG, Cooper WT. 2010. Comprehensive characterization of marine dissolved organic matter by fourier transform ion cyclotron resonance mass spectrometry with electrospray and atmospheric pressure photoionization. Rapid Commun Mass Spectrom. 24(5):643-650.
Daling PS, Faksness LG, Hansen AB, Stout SA. 2002. Improved and standardized methodology for oil spill fingerprinting. Environ Forensics. 3(3-4):263-278.
Damasceno FC, Gruber LDA, Geller AM, de Campos MCV, Gomes AO, Guimaraes RCL, Peres VF, Jacquesg RA, Caramao EB. 2014. Characterization of naphthenic acids using mass spectroscopy and chromatographic techniques: Study of technical mixtures. Anal Methods-Uk. 6(3):807-816.
de Carvalho Rocha WF, Schantz MM, Sheen DA, Chu PM, Lippa KA. 2017. Unsupervised classification of petroleum certified reference materials and other fuels by chemometric analysis of gas chromatography-mass spectrometry data. Fuel (Lond). 197:248-258.
Dimitrov SD, Georgieva DG, Pavlov TS, Karakolev YH, Karamertzanis PG, Rasenberg M, Mekenyan OG. 2015. Uvcb substances: Methodology for structural description and application to fate and hazard assessment. Environ Toxicol Chem. 34(11):2450-2462.
Dodds JN, Baker ES. 2019. Ion mobility spectrometry: Fundamental concepts, instrumentation, applications, and the road ahead. Journal of the American Society for Mass Spectrometry. 30(11):2185-2195.

Dodds JN, Hopkins ZR, Knappe DRU, Baker ES. 2020. Rapid characterization of per- and polyfluoroalkyl substances (pfas) by ion mobility spectrometry-mass spectrometry (ims-ms). Anal Chem. 92(6):4427-4435.
Dong X, Wang F, Fan X, Zhao YP, Wei XY, Wang RY, Ma FY, Liu JM, Li B. 2019. Evaluation of elemental composition obtained by using mass spectrometer and elemental analyzer: A case study on model compound mixtures and a coal-derived liquid. Fuel. 245:392-397.
Dunn OJ. Multiple comparison among means. J Am Stat Assoc. (56):52-64.
ECHA. 2008. Guidance on information requirements and chemical safety assessment. Chapter r.6: Qsars and grouping of chemicals.
ECHA. 2016. Preparation of an inventory of substances suspected to meet reach annex iii criteria. Helsinki, Finland: European Chemicals Agency.
ECHA. 2017a. Guidance for identification and naming of substances under reach and clp. Helsinki, Finland: European Chemical Agency.
ECHA. 2017b. Guidance on requirements for substances in articles. Helsinki, Finland: European Chemicals Agency.
ECHA. 2020. Testing proposal decision on substance ec 295-332-8 "extracts (petroleum), deasphalted vacuum residue solvent". Helsinki, Finland: European Chemicals Agency.
El-Sabagh S, El-Naggar A, El Nady M, Ebiad M, Rashad A, Abdullah E. 2018. Distribution of triterpanes and steranes biomarkers as indication of organic matters input and depositional environments of crude oils of oilfields in gulf of suez, egypt. Egyptian journal of petroleum. 27(4):969-977.
Elbaz AM, Gani A, Hourani N, Emwas AH, Sarathy SM, Roberts WL. 2015. Tg/dtg, ft-icr mass spectrometry, and nmr spectroscopy study of heavy fuel oil. Energy & Fuels. 29(12):7825-7835.
EPA US. 1995. Toxic substances control act inventory representation for chemical substances of unknown or variable composition, complex reaction products and biological materials: Uvcb substances.
European Commission. 2021. Commission regulation (eu) 2021/979 of 17 june 2021 amending annexes vii to xi to regulation (ec) no 1907/2006 of the european parliament and of the council concerning the registration, evaluation, authorisation and restriction of chemicals (reach). Brussels, Belgium: Official Journal of the European Union.
Everitt B. 1980. Cluster analysis. Qual Quant. 14(1):75-100.
Faksness LG, Daling PS, Hansen AB. 2002. Round robin study—oil spill identification. Environ Forensics. 3(3-4):279-291.
Farenc M, Corilo YE, Lalli PM, Riches E, Rodgers RP, Afonso C, Giusti P. 2016. Comparison of atmospheric pressure ionization for the analysis of heavy petroleum fractions with ion mobility-mass spectrometry. Energ Fuel. 30:8896-8903.
Fathalla EM, Andersson JT. 2011. Products of polycyclic aromatic sulfur heterocycles in oil spill photodegradation. Environ Toxicol Chem. 30(9):2004-2012.

Fernandez-Lima FA, Becker C, McKenna AM, Rodgers RP, Marshall AG, Russell DH. 2009. Petroleum crude oil characterization by ims-ms and fticr ms. Analytical chemistry. 81(24):9941-9947.
Forsberg ND, O'Connell SG, Allan SE, Anderson KA. 2014. Passive sampling coupled to ultraviolet irradiation: A useful analytical approach for studying oxygenated polycyclic aromatic hydrocarbon formation in bioavailable mixtures. Environ Toxicol Chem. 33(1):177-181.
Fowlkes EB, Mallows CL. 1983. A method for comparing two hierarchical clusterings. J Am Stat Assoc. 78(383):553-569.
Frenzel M, Scarlett A, Rowland SJ, Galloway TS, Burton SK, Lappin-Scott HM, Booth AM. 2010. Complications with remediation strategies involving the biodegradation and detoxification of recalcitrant contaminant aromatic hydrocarbons. Sci Total Environ. 408(19):4093-4101.
Frysinger GS, Gaines RB, Xu L, Reddy CM. 2003. Resolving the unresolved complex mixture in petroleum-contaminated sediments. Environmental science & technology. 37(8):1653-1662.
Gabelica V, Shvartsburg AA, Afonso C, Barran P, Benesch JLP, Bleiholder C, Bowers MT, Bilbao A, Bush MF, Campbell JL et al. 2019. Recommendations for reporting ion mobility mass spectrometry measurements. Mass Spectrom Rev. 38(3):291-320.
Gabetti E, Sgorbini B, Stilo F, Bicchi C, Rubiolo P, Chialva F, Reichenbach SE, Bongiovanni V, Cordero C, Cavallero A. 2021. Chemical fingerprinting strategies based on comprehensive two-dimensional gas chromatography combined with gas chromatography-olfactometry to capture the unique signature of piemonte peppermint essential oil (mentha x piperita var italo-mitcham). J Chromatogr A. 1645:462101.
Gaines RB, Frysinger GS, Hendrick-Smith MS, Stuart JD. 1999. Oil spill source identification by comprehensive two-dimensional gas chromatography. Environmental science & technology. 33(12):2106-2112.
Giles HN. 2016. Crude oil analysis: History and development of test methods from 1854 to 2016. Mater Perform Charact. 5:1-169.
Gobas F, Mayer P, Parkerton TF, Burgess RM, van de Meent D, Gouin T. 2018. A chemical activity approach to exposure and risk assessment of chemicals: Focus articles are part of a regular series intended to sharpen understanding of current and emerging topics of interest to the scientific community. Environ Toxicol Chem. 37(5):1235-1251.
Griffiths MT, Da Campo R, O'Connor PB, Barrow MP. 2014. Throwing light on petroleum: Simulated exposure of crude oil to sunlight and characterization using atmospheric pressure photoionization fourier transform ion cyclotron resonance mass spectrometry. Anal Chem. 86(1):527-534.
Grimm FA, Iwata Y, Sirenko O, Chappell GA, Wright FA, Reif DM, Braisted J, Gerhold DL, Yeakley JM, Shepard P et al. 2016. A chemical-biological similarity-based grouping of complex substances as a prototype approach for evaluating chemical alternatives. Green Chem. 18(16):4407-4419.

Grimm FA, Russell WK, Luo YS, Iwata Y, Chiu WA, Roy T, Boogaard PJ, Ketelslegers HB, Rusyn I. 2017. Grouping of petroleum substances as example uvcbs by ion mobility-mass spectrometry to enable chemical composition-based read-across. Environmental science & technology. 51(12):7197-7207.

Hpv challenge overview. 2017. API; [accessed]. https://www.petroleumhpv.org/hpv-challenge.

Guillemant J, Albrieux F, Lacoue-Negre M, Pereira de Oliveira L, Joly JF, Duponchel L. 2019. Chemometric exploration of appi(+)-ft-icr ms data sets for a comprehensive study of aromatic sulfur compounds in gas oils. Anal Chem. 91(18):11785-11793.

Hansen AB, Daling PS, Faksness L, Sorheim KR, Kienhuis P, Duus R. 2007. Emerging cen methodology for oil spill identification. In: Wang Z, Stout SA, editors. Chemical fingerprinting of spilled or discharged petroleum — methods and factors affecting petroleum fingerprints in the environment. Cambridge, MA: Academic Press. p. 229-256.

Hantao LW, Aleme HG, Pedroso MP, Sabin GP, Poppi RJ, Augusto F. 2012. Multivariate curve resolution combined with gas chromatography to enhance analytical separation in complex samples: A review. Anal Chim Acta. 731:11-23.

Hawkes JA, D'Andrilli J, Agar JN, Barrow MP, Berg SM, Catalán N, Chen H, Chu RK, Cole RB, Dittmar T et al. 2020. An international laboratory comparison of dissolved organic matter composition by high resolution mass spectrometry: Are we getting the same answer? Limnol Oceanogr Methods. 18(6):235-258.

Headly JV, Peru KM, Barrow MP. 2015. Advances in mass spectrometric characterization of naphthenic acids fraction compounds in oil sands environmental samples and crude oil - a review. Mass Spectrom Rev. 35:311-328.

Hedgpeth BM, Redman AD, Alyea RA, Letinski DJ, Connelly MJ, Butler JD, Zhou H, Lampi MA. 2019. Analysis of sublethal toxicity in developing zebrafish embryos exposed to a range of petroleum substances. Environ Toxicol Chem. 38(6):1302-1312.

Hoskins JN, Trimpin S, Grayson SM. 2011. Architectural differentiation of linear and cyclic polymeric isomers by ion mobility spectrometry-mass spectrometry. Macromolecules. 44(17):6915-6918.

Hosseini SH, Sachsenhofer. 2021. Characterization of crude oils derived from carbonate and siliciclastic source rocks using fticr-ms. Org Geochem. 159.

House JS, Grimm FA, Klaren WD, Dalzell A, Kuchi S, Zhang SD, Lenz K, Boogaard PJ, Ketelslegers HB, Gant TW et al. 2021. Grouping of uvcb substances with new approach methodologies (nams) data. ALTEX. 38(1):123-137.

Hsu CS, Hendrickson CL, Rodgers RP, McKenna AM, Marshall AG. 2011. Petroleomics: Advanced molecular probe for petroleum heavy ends. J Mass Spectrom. 46(4):337-343.

Hsu CS, Shi Q. 2013. Prospects for petroleum mass spectrometry and chromatography. Sci China Chem. 56:833-839.

Hughey CA, Hendrickson CL, Rodgers RP, Marshall AG, Qian K. 2001. Kendrick mass defect spectrum: A compact visual analysis for ultrahigh-resolution broadband mass spectra. Anal Chem. 73(19):4676-4681.
Hur M, Yeo I, Park E, Kim YH, Yoo J, Kim EJ, No MH, Koh J, Kim S. 2010. Combination of statistical methods and fourier transform ion cyclotron resonance mass spectrometry for more comprehensive, molecular-level interpretations of petroleum samples. Anal Chem. 82(1):211-218.
Ibrahim YM, Baker ES, Danielson WF, 3rd, Norheim RV, Prior DC, Anderson GA, Belov ME, Smith RD. 2015. Development of a new ion mobility (quadrupole) time-of-flight mass spectrometer. Int J Mass Spectrom. 377:655-662.
Ibrahim YM, Garimella SV, Prost SA, Wojcik R, Norheim RV, Baker ES, Rusyn I, Smith RD. 2016. Development of an ion mobility spectrometry-orbitrap mass spectrometer platform. Anal Chem. 88(24):12152-12160.
Islam A, Cho Y, Ahmed A, Kim S. 2012. Data interpretation methods for petroleomics. Mass Spectrometry Letters. 3(3):63-67.
Jaggi A, Radovic JR, Snowdon LR, Larter SR, Oldenburg TBP. 2019. Composition of the dissolved organic matter produced during in situ burning of spilled oil. Org Geochem. 138.
Juahir H, Ismail A, Mohamed SB, Toriman ME, Kassim AM, Zain SM, Ahmad WKW, Wah WK, Zali MA, Retnam A et al. 2017. Improving oil classification quality from oil spill fingerprint beyond six sigma approach. Mar Pollut Bull. 120(1-2):322-332.
Kaiser MJ. 2017. A review of refinery complexity applications. Pet Sci. 14:167-194.
Kauppila TJ, Kuuranne T, Meurer EC, Eberlin MN, Kotiaho T, Kostiainen R. 2002. Atmospheric pressure photoionization mass spectrometry. Ionization mechanism and the effect of solvent on the ionization of naphthalenes. Anal Chem. 74(21):5470-5479.
Kendrick E. 1963. A mass scale based on ch2 = 14.0000 for high resolution mass spectrometry of organic compounds. Anal Chem. 35(13):2146-2154.
Kennicutt II MC, McDonald TJ, Comet PA, Denoux GJ, Brooks JM. 1992. The origins of petroleum in the northern gulf of mexico. Geochim Cosmochim Acta. 56(3):1259-1280.
Keramea P, Spanoudaki K, Zodiatis G, Gikas G, Sylaios G. 2021. Oil spill modeling: A critical review on current trends, perspectives, and challenges. J Mar Sci Eng. 9(2).
Kim D, Jin JM, Cho Y, Kim EH, Cheong HK, kim YH, Kim S. 2015. Combination of ring type hplc separation, ultrahigh-resolution mass spectrometry, and high field nmr for comprehensive characterization of crude oil compositions. Fuel. 157:48-55.
Kim D, Jung JH, Ha SY, An JG, Shankar R, Kwon JH, Yim UH, Kim SH. 2019. Molecular level determination of water accommodated fraction with embryonic developmental toxicity generated by photooxidation of spilled oil. Chemosphere. 237:124346.

Kim S, Kramer RW, Hatcher PG. 2003. Graphical method for analysis of ultrahigh-resolution broadband mass spectra of natural organic matter, the van krevelen diagram. Anal Chem. 75(20):5336-5344.
King SM, Leaf PA, Olson AC, Ray PZ, Tarr MA. 2014. Photolytic and photocatalytic degradation of surface oil from the deepwater horizon spill. Chemosphere. 95:415-422.
Koch BP, Dittmar T, Witt M, Kattner G. 2007. Fundamentals of molecular formula assignment to ultrahigh resolution mass data of natural organic matter. Anal Chem. 79(4):1758-1763.
Kooijman PC, Nagornov KO, Kozhinov AN, Kilgour DPA, Tsybin YO, Heeren RMA, Ellis SR. 2019. Increased throughput and ultra-high mass resolution in desi ft-icr ms imaging through new-generation external data acquisition system and advanced data processing approaches. Sci Rep. 9(1):8.
Laffon B, Pasaro E, Valdiglesias V. 2016. Effects of exposure to oil spills on human health: Updated review. J Toxicol Environ Health B Crit Rev. 19(3-4):105-128.
Lalli PM, Corilo YE, Rowland SJ, Marshall AG, Rodgers RP. 2015. Isomeric separation and structural characterization of acids in petroleum by ion mobility mass spectrometry. Energ Fuel. 29:3626-3633.
Lalli PM, Jarvis JM, Marshall AG, Rodgers RP. 2017. Functional isomers in petroleum emulsion interfacial material revealed by ion mobility mass spectrometry and collision-induced dissociation. Energy & Fuels. 31(1):311-318.
Letinski D, Parkerton T, Redman A, Manning R, Bragin G, Febbo E, Palandro D, Nedwed T. 2014. Use of passive samplers for improving oil toxicity and spill effects assessment. Mar Pollut Bull. 86(1-2):274-282.
Li Y, Wu B, He C, Nie F, Shi Q. 2022. Comprehensive chemical characterization of dissolved organic matter in typical point-source refinery wastewaters. Chemosphere. 286(Pt 1):131617.
Liu Y, Huang H, Liu Q, Xu X, Cheng H. 2020. The acid and neutral nitrogen compounds characterized by negative esi orbitrap ms in a heavy oil before and after oxidation. Fuel. 277.
Llompart M, Celeiro M, Garcia-Jares C, Dagnac T. 2019. Environmental applications of solid-phase microextraction. Trac-Trends in Analytical Chemistry. 112:1-12.
Lübeck JS, Alexandrino GL, Christensen JH. 2020. Gc × gc–hrms nontarget fingerprinting of organic micropollutants in urban freshwater sediments. Environ Sci Eur. 32(1):78.
Luna N, Zelong L, Songbai T. 2014. Identification and characterization of sulfur compounds in straight-run diesel using comprehensive two-dimensional gas chromatography coupled to time-of-flight mass spectrometry. China Petrol Proc Petrochem Techn. 16(3):10-18.
Luo YS, Aly NA, McCord J, Strynar MJ, Chiu WA, Dodds JN, Baker ES, Rusyn I. 2020a. Rapid characterization of emerging per- and polyfluoroalkyl substances in aqueous film-forming foams using ion mobility spectrometry-mass spectrometry. Environmental science & technology. 54(23):15024-15034.

Luo YS, Ferguson KC, Rusyn I, Chiu WA. 2020b. In vitro bioavailability of the hydrocarbon fractions of dimethyl sulfoxide extracts of petroleum substances. Toxicol Sci. 174(2):168-177.
Mahmoud MY, Dabek-Zlotorzynska E. 2018. Investigation of isomeric structures in a commercial mixture of naphthenic acids using ultrahigh pressure liquid chromatography coupled to hybrid traveling wave ion mobility-time of flight mass spectrometry. J Chromatogr A. 1572:90-99.
Maillard JF, Le Maitre J, Ruger CP, Ridgeway M, Thompson CJ, Paupy B, Hubert-Roux M, Park M, Afonso C, Giusti P. 2021. Structural analysis of petroporphyrins from asphaltene by trapped ion mobility coupled with fourier transform ion cyclotron resonance mass spectrometry. Analyst. 146(13):4161-4171.
Maki H, Sasaki T, Harayama S. 2001. Photo-oxidation of biodegraded crude oil and toxicity of the photo-oxidized products. Chemosphere. 44(5):1145-1151.
Marshall AG, Blakney GT, Beu SC, Hendrickson CL, McKenna AM, Purcell JM, Rodgers RP, Xian F. 2010. Petroleomics: A test bed for ultra-high-resolution fourier transform ion cyclotron resonance mass spectrometry. Eur J Mass Spectrom (Chichester). 16(3):367-371.
Marshall AG, Rodgers RP. 2004. Petroleomics: The next grand challenge for chemical analysis. Acc Chem Res. 37(1):53-59.
Marshall AG, Rodgers RP. 2008. Petroleomics: Chemistry of the underworld. Proc Natl Acad Sci U S A. 105(47):18090-18095.
Marvel SW, To K, Grimm FA, Wright FA, Rusyn I, Reif DM. 2018. Toxpi graphical user interface 2.0: Dynamic exploration, visualization, and sharing of integrated data models. BMC Bioinf. 19(1):80.
McDonald TJ, Kennicutt II MC. 1992. Fractionation of crude oils by hplc and quantitative determination of aliphatic and aromatic biological markers by gc/ms-sim. LC-GC. 10(12):935-938.
McKee RH, Adenuga MD, Carrillo JC. 2015. Characterization of the toxicological hazards of hydrocarbon solvents. Critical reviews in toxicology. 45(4):273-365.
McKee RH, Medeiros AM, Daughtrey WC. 2005. A proposed methodology for setting occupational exposure limits for hydrocarbon solvents. J Occup Environ Hyg. 2(10):524-542.
McKee RH, White R. 2014. The mammalian toxicological hazards of petroleum-derived substances: An overview of the petroleum industry response to the high production volume challenge program. Int J Toxicol. 33(1 Suppl):4S-16S.
McKenna AM, Chen H, Weisbrod CR, Blakney GT. 2021. Molecular comparison of solid-phase extraction and liquid/liquid extraction of water-soluble petroleum compounds produced through photodegradation and biodegradation by ft-icr mass spectrometry. Anal Chem. 93(10):4611-4618.
McKenna AM, Nelson RK, Reddy CM, Savory JJ, Kaiser NK, Fitzsimmons JE, Marshall AG, Rodgers RP. 2013. Expansion of the analytical window for oil spill characterization by ultrahigh resolution mass spectrometry: Beyond gas chromatography. Environmental science & technology. 47(13):7530-7539.

McLafferty F.W., F. T. 1993. Interpretation of mass spectra. Mill Valley, Californa: University Science Books.
Mennito AS, Qian K. 2013. Characterization of heavy petroleum saturates by laser desorption silver cationization and fourier transform ion cyclotron resonance mass spectrometry. Energy & Fuels. 27(12):7348-7353.
Metz TO, Baker ES, Schymanski EL, Renslow RS, Thomas DG, Causon TJ, Webb IK, Hann S, Smith RD, Teeguarden JG. 2017. Integrating ion mobility spectrometry into mass spectrometry-based exposome measurements: What can it add and how far can it go? Bioanalysis. 9(1):81-98.
Miles SM, Asiedy E, Balaberda AI, Ulrich AC. 2020. Oil sands process affected water sourced trichoderma harzianum demonstrates capacity for mycoremediation of naphthenic acid fraction compounds. Chemosphere. 258.
Muller H, Alawani NA, Adam FM. 2020. Innate sulfur compounds as an internal standard for determining vacuum gas oil compositions by appi ft-icr ms. Energ Fuel. 34(7):8260-8273.
NASEM. 2020. The use of dispersants in marine oil spill response. Washington, DC: The National Academies Press.
National Research Council. 1999. Spills of non-floating oils: Risk and response. Washington, DC.
National Research Council. 2003. Oil in the sea iii: Inputs, fates, and effects. The National Academies Press Washington, DC. p. 265.
National Research Council. 2014. Responding to oil spills in the u.S. Arctic marine environment. Washington, DC.
Ngo HI, Hoh E, Foglia TA. 2012. Improved synthesis and characterization of saturated branched-chain fatty acid isomers. Eur J Lipid Sci Technol. 114:213-221.
Nicodem DE, Fernandes MCZ, Guedes CLB, Correa RJ. 1997. Photochemical processes and the environmental impact of petroleum spills. Biogeochem. 39:121-138.
Nicodem DE, Guedes CLB, Fernandes MCZ, Severino D, Correa RJ, Coutinho MC, Silva J. 2001. Photochemistry of petroleum. Prog React Kinet Mech. 26:219-238.
Niyonsaba E, Manheim JM, Yerabolu R, Kenttamaa HI. 2019. Recent advances in petroleum analysis by mass spectrometry. Anal Chem. 91(1):156-177.
O'Reilly KT, Mohler RE, Zemo DA, Ahn S, Magaw RI, Espino Devine C. 2019. Oxygen-containing compounds identified in groundwater from fuel release sites using gcxgc-tof-ms. Groundwater Monit Remediat. 39(4):32-40.
OECD. 2014. Guidance on grouping of chemicals, second edition. Paris, France.
OECD. 2019. Guidance document on aqueous-phase aquatic toxicity testing of difficult test chemicals. Paris, France.
Oldenburg TBP, Brown M, Bennet B, Larter SR. 2014. The impact of thermal maturity level on the composition of crude oils, assessed using ultra-high resolution mass spectrometry. Org Geochem. 75:151-168.
Oldenburg TBP, Jones M, Huang H, Bennet B, Shafiee NS, Head I, Larter SR. 2017. The controls on the composition of biodegraded oils in the deep subsurface – part 4. Destruction and production of high molecular weight non-hydrocarbon species

and destruction of aromatic hydrocarbons during progressive in-reservoir biodegradation. Org Geochem. 114:57-80.
Onel M, Beykal B, Ferguson K, Chiu WA, McDonald TJ, Zhou L, House JS, Wright FA, Sheen DA, Rusyn I et al. 2019. Grouping of complex substances using analytical chemistry data: A framework for quantitative evaluation and visualization. PloS one. 14(10):e0223517.
Orrego- Ruíz JA. 2018. Finding a relationship between the composition and the emulsifying character of asphaltenes through fticr-ms. Cienc Tecn Fut. 8(1):45-52.
Palacio Lozano DC, Chacon-Patiño ML, Gomez-Escudero A, Barrow MP. 2019a. Chapter 32 | mass spectrometry in the petroleum industry," in fuels and lubricants handbook: Technology, properties, performance, and testing. MNL37-2ND-EB Fuels and Lubricants Handbook: Technology, Properties, Performance, and Testing. 2.
Palacio Lozano DC, Gavard R, Arenas-Diaz JP, Thomas MJ, Stranz DD, Mejia-Ospino E, Guzman A, Spencer SEF, Rossell D, Barrow MP. 2019b. Pushing the analytical limits: New insights into complex mixtures using mass spectra segments of constant ultrahigh resolving power. Chem Sci. 10(29):6966-6978.
Palacio Lozano DC, Thomas MJ, Jones HE, Barrow MP. 2020. Petroleomics: Tools, challenges, and developments. Annu Rev Anal Chem (Palo Alto Calif). 13(1):405-430.
Pang XQ, Jia CZ, Wang WY. 2015. Petroleum geology features and research developments of hydrocarbon accumulation in deep petroliferous basins. Petroleum Science. 12(1):1-53.
Payne JR, Phillips CR. 1985. Photochemistry of petroleum in water. Environ Sci Technol. 19(7):569-579.
Pelletier MC, Burgess RM, Ho KT, Kuhn A, McKinney RA, Ryba SA. 1997. Phototoxicity of individual polycyclic aromatic hydrocarbons and petroleum to marine invertebrate larvae and juveniles. Environmental Toxicology and Chemistry. 16(10):2190-2199.
Peters KE, Walters CC, Moldowan JM. 2005. The biomarker guide. Cambridge, United Kingdom: Cambridge University Press.
Ponthus J, Riches E. 2013. Evaluating the multiple benefits offered by ion mobility-mass spectrometry in oil and petroleum analysis. Int J Ion Mobil Spec. 16(2):95-103.
Prince RC, Walters CC. 2022. Modern analytical techniques are improving our ability to follow the fate of spilled oil in the environment. Current Opinion in Chemical Engineering. 36:100787.
Purcell JM, Hendrickson CL, Rodgers RP, Marshall AG. 2007a. Atmospheric pressure photoionization proton transfer for complex organic mixtures investigated by fourier transform ion cyclotron resonance mass spectrometry. J Am Soc Mass Spectrom. 18(9):1682-1689.
Purcell JM, Rodgers RP, Hendrickson CL, Marshall AG. 2007b. Speciation of nitrogen containing aromatics by atmospheric pressure photoionization or electrospray

ionization fourier transform ion cyclotron resonance mass spectrometry. J Am Soc Mass Spectrom. 18(7):1265-1273.
Pusyn T, Leszczynski J, Cronin MTD. 2009. Recent advances in qsar studies. Liverpool, UK: Springer.
Qian K, Wang FC. 2019. Compositional analysis of heavy petroleum distillates by comprehensive two-dimensional gas chromatography, field ionization and high-resolution mass spectrometry. J Am Soc Mass Spectrom. 30(12):2785-2794.
Radovic JR, Rial D, Lyons BP, Harman C, Vinas L, Beiras R, Readman JW, Thomas KV, Bayona JM. 2012. Post-incident monitoring to evaluate environmental damage from shipping incidents: Chemical and biological assessments. J Environ Manage. 109:136-153.
Rasmussen K, Pettauer D, Vollmer G, Davis J. 1998. Compilation of einecs: Description and definitions used for uvcb substances: Complex reaction products, plant products, (post-reacted) naturally occurring substances, micro-organisms, petroleum products, soaps and detergents, and metallic compounds. Toxicol Environ Chem. 69:403-416.
Ray PZ, Chen H, Podgorski DC, McKenna AM, Tarr MA. 2014. Sunlight creates oxygenated species in water-soluble fractions of deepwater horizon oil. J Hazard Mater. 280:636-643.
Reddy CM, Quinn JG. 1999. Gc-ms analysis of total petroleum hydrocarbons and polycyclic aromatic hydrocarbons in seawater samples after the north cape oil spill. Marine Pollution Bulletin. 38(2):126-135.
Redman AD, Butler JD, Letinski DJ, Di Toro DM, Leon Paumen M, Parkerton TF. 2018a. Technical basis for using passive sampling as a biomimetic extraction procedure to assess bioavailability and predict toxicity of petroleum substances. Chemosphere. 199:585-594.
Redman AD, Parkerton TF, Butler JD, Letinski DJ, Frank RA, Hewitt LM, Bartlett AJ, Gillis PL, Marentette JR, Parrott JL et al. 2018b. Application of the target lipid model and passive samplers to characterize the toxicity of bioavailable organics in oil sands process-affected water. Environmental science & technology. 52(14):8039-8049.
Ristic ND, Djokic MR, Delbeke E, Gonzalez-Quiroga A, Stevens CV, Van Geem KM, Marin GB. 2018. Compositional characterization of pyrolysis fuel oil from naphtha and vacuum gas oil. Energ Fuel. 32(2):1276-1286.
Rocha YDS, Pereira RCL, Graciano J, Filho M. 2018. Geochemical characterization of lacustrine and marine oils from off-shore brazilian sedimentary basins using negative-ion electrospray fourier transform ion cyclotron resonance mass spectrometry (esi fticr-ms). Org Geochem. 124:29-45.
Rodgers RP, McKenna AM. 2011. Petroleum analysis. Anal Chem. 83(12):4665-4687.
Roman-Hubers AT, Cordova AC, Aly NA, McDonald TJ, Lloyd DT, Wright FA, Baker ES, Chiu WA, Rusyn I. 2021a. Data processing workflow to identify structurally related compounds in petroleum substances using ion mobility spectrometry-mass spectrometry. Energ Fuel. 35(13):10529-10539.

Roman-Hubers AT, McDonald TJ, Baker ES, Chiu WA, Rusyn I. 2021b. A comparative analysis of analytical techniques for rapid oil spill identification. Environ Toxicol Chem. 40:1034-1049.
Rowland SJ, West CE, Scarlett AG, Jones D. 2011. Identification of individual acids in a commercial sample of naphthenic acids from petroleum by two-dimensional comprehensive gas chromatography/mass spectrometry. Rapid Commun Mass Spectrom. 25:1741-1751.
Ruger CP, Le Maitre J, Maillard J, Riches E, Palmer M, Afonso C, Giusti P. 2021. Exploring complex mixtures by cyclic ion mobility high-resolution mass spectrometry: Application toward petroleum. Anal Chem. 93(14):5872-5881.
Salvito D, Fernandez M, Jenner K, Lyon DY, de Knecht J, Mayer P, MacLeod M, Eisenreich K, Leonards P, Cesnaitis R et al. 2020. Improving the environmental risk assessment of substances of unknown or variable composition, complex reaction products, or biological materials. Environ Toxicol Chem. 39(11):2097-2108.
Santos JM, Galaverna RD, Pudenzi MA, Schmidt EM, Sanders NL, Kurulugama RT, Mordehai A, Stafford GC, Wisniewski A, Eberlin MN. 2015. Petroleomics by ion mobility mass spectrometry: Resolution and characterization of contaminants and additives in crude oils and petrofuels. Anal Methods. 7(11):4450-4463.
Scarlett A, Rowland SJ, Galloway TS, Lewis AC, Booth AM. 2008. Chronic sublethal effects associated with branched alkylbenzenes bioaccumulated by mussels. Environ Toxicol Chem. 27(3):561-567.
Shankar R, Shim WJ, An JG, Yim UH. 2015. A practical review on photooxidation of crude oil: Laboratory lamp setup and factors affecting it. Water Res. 68:304-315.
Silva RC, Snowdon LR, Huang H, Larter S. 2021. The dating of petroleum fluid residence time in subsurface reservoirs. Part 2: Tracking effects of radiolysis on crude oil by comprehensive molecular analysis. Org Geochem. 152.
Silva RC, Yim C, Radovic JR, Brown M, Weerawardhena P, Huang H, Showdon LR, Oldenburg TBP, Larter SR. 2020. Mechanistic insights into sulfur rich oil formation, relevant to geological carbon storage routes. A study using (+) appi fticr-ms analysis. Org Geochem. 147.
Snyder K, Mladenov N, Richardot W, Dodder N, Nour A, Campbell C, Hoh E. 2021. Persistence and photochemical transformation of water soluble constituents from industrial crude oil and natural seep oil in seawater. Mar Pollut Bull. 165:112049.
Staš M, Chudoba J, Kubička D, Blažek J, Pospíšil M. 2017. Petroleomic characterization of pyrolysis bio-oils: A review. Energy & Fuels. 31(10):10283-10299.
Stewart GJ, Nelson BS, Acton WJF, Vaughan AR, Farren NJ, Hopkins JR, Ward MW, Swift SJ, Arya R, Mondal A et al. 2021. Emissions of intermediate-volatility and semi-volatile organic compounds from domestic fuels used in delhi, india. Atmos Chem Phys. 21(4):2407-2426.
Stout SA, Wang Z. 2007. Chemical fingerprinting of spilled or discharged petroleum — methods and factors affecting petroleum fingerprints in the environment. In: Wang Z, Stout SA, editors. Oil spill environmental forensics: Fingerprinting and source identification. Cambridge, MA: Academic Press. p. 1-53.

Stout SA, Wang Z. 2016. Standard handbook oil spill environmental forensics: Fingerprinting and source identification. Academic Press.
Stow SM, Causon TJ, Zheng X, Kurulugama RT, Mairinger T, May JC, Rennie EE, Baker ES, Smith RD, McLean JA et al. 2017. An interlaboratory evaluation of drift tube ion mobility-mass spectrometry collision cross section measurements. Anal Chem. 89(17):9048-9055.
Sydnes LK, Hemmingsen TH, Skare S, Hansen SH, Falk-Petersen IB, Loenning S, Oestgaard K. 1985. Seasonal variations in weathering and toxicity of crude oil on seawater under arctic conditions. Environmental science & technology. 19(11):1076-1081.
Tarr MA, Zito P, Overton EB, Olsen GM, Puspa LA, Reddy CM. 2016. Weathering of oil spilled in the marine environment. Oceanography. 29:126-135.
Terra LA, Filgueiras PR, Tose LV, Romao W, de Souza DD, de Castro EV, de Oliveira MS, Dias JC, Poppi RJ. 2014. Petroleomics by electrospray ionization ft-icr mass spectrometry coupled to partial least squares with variable selection methods: Prediction of the total acid number of crude oils. Analyst. 139(19):4908-4916.
U.S. EPA. 1978. Toxic substance control act (tsca) pl 94-469 candidate list of chemical substances addendum 1 generic terms covering petroleum refinery process streams. Washignton, D.C.: US Environmental Protection Agency,.
US EPA. 1996. Method 8051b: Non-halogenated organics using gc-fid. Washington, DC: US Environmental Protection Agency.
US EPA. 2014. Method 8270e (sw-846): Semivolatile organic compounds by gas chromatography/ mass spectrometry (gc/ms). Washington, DC: US Environmental Protection Agency.
Van De Weghe H, Vanermen G, Gemoets J, Lookman R, Bertels D. 2006. Application of comprehensive two-dimensional gas chromatography for the assessment of oil contaminated soils. J Chromatogr A. 1137(1):91-100.
van Krevelen DW. 1950. Graphical-statistical method for the study of structure and reaction processes of coal. Fuel. 29:269-284.
Van Krevelen DW. 1984. Organic geochemistry - old and new. Org Geochem. 6:1-10.
Vanini G, Barra TA, Souza LM, Madeira NCL, Gomes AO, Romao W, Azevedo DA. 2020. Characterization of nonvolatile polar compounds from brazilian oils by electrospray ionization with ft-icr ms and orbitrap-ms. Fuel. 282.
Vaughan PP, Wilson T, Kamerman R, Hagy ME, McKenna A, Chen H, Jeffrey WH. 2016. Photochemical changes in water accommodated fractions of mc252 and surrogate oil created during solar exposure as determined by ft-icr ms. Mar Pollut Bull. 104(1-2):262-268.
Ventura GT, Hall GJ, Nelson RK, Frysinger GS, Raghuraman B, Pomerantz AE, Mullins OC, Reddy CM. 2011. Analysis of petroleum compositional similarity using multiway principal components analysis (mpca) with comprehensive two-dimensional gas chromatographic data. J Chromatogr A. 1218(18):2584-2592.
Vozka P, Modereger BA, Park AC, Zhang WTJ, Trice RW, Kenttamaa HI, Kilaz G. 2019. Jet fuel density via gc x gc-fid. Fuel. 235:1052-1060.

Walters CC, Wang FC, Qian K, Wu C, Mennito AS, Wei Z. 2015. Petroleum alteration by thermochemical sulfate reduction – a comprehensive molecular study of aromatic hydrocarbons and polar compounds. Geochim Cosmochim Acta. 153:37-71.
Wang FC, Qian K, Green LA. 2005. Gcxms of diesel: A two-dimensional separation approach. Anal Chem. 77(9):2777-2785.
Wang X, Ji Y, Shi Q, Zhang Y, He C, Wang Q, Guo S, Chen C. 2020. Characterization of wastewater effluent organic matter with different solid phase extraction sorbents. Chemosphere. 257:127235.
Wang Z, Fingas MF. 2003a. Development of oil hydrocarbon fingerprinting and identification techniques. Mar Pollut Bull. 47(9-12):423-452.
Wang Z, Stout S. 2010. Oil spill environmental forensics: Fingerprinting and source identification. Elsevier.
Wang Z, Stout SA, Fingas M. 2006. Forensic fingerprinting of biomarkers for oil spill characterization and source identification. Environ Forensics. 7(2):105-146.
Wang Z, Yang C, Yang Z, Sun J, Hollebone B, Brown C, Landriault M. 2011. Forensic fingerprinting and source identification of the 2009 sarnia (ontario) oil spill. J Environ Monit. 13(11):3004-3017.
Wang ZD, Fingas M. 2003b. Fate and identification of spilled oils and petroleum products in the environment by gc-ms and gc-fid. Energy Sources. 25(6):491-508.
Ward CP, Overton EB. 2020. How the 2010 deepwater horizon spill reshaped our understanding of crude oil photochemical weathering at sea: A past, present, and future perspective. Environ Sci Process Impacts. 22(5):1125-1138.
Weng N, Wan S, Wang H, Zhang S, Zhu G, Liu J, Cai D, Yang Y. 2015. Insight into unresolved complex mixtures of aromatic hydrocarbons in heavy oil via two-dimensional gas chromatography coupled with time-of-flight mass spectrometry analysis. J Chromatogr A. 1398:94-107.
White HK, Morrison AE, Dhoonmoon C, Caballero-Gomez H, Luu M, Samuels C, Marx CT, Michel APM. 2020. Identification of persistent oil residues in prince william sound, alaska using rapid spectroscopic techniques. Mar Pollut Bull. 161(Pt B):111718.
Williams ES, Panko J, Paustenbach DJ. 2009. The european union's reach regulation: A review of its history and requirements. Critical reviews in toxicology. 39(7):553-575.
Wozniak AS, Prem PM, Obeid W, Waggoner DC, Quigg A, Xu C, Santschi PH, Schwehr KA, Hatcher PG. 2019. Rapid degradation of oil in mesocosm simulations of marine oil snow events. Environmental science & technology. 53(7):3441-3450.
Xian F, Hendrickson CL, Marshall AG. 2012. High resolution mass spectrometry. Anal Chem. 84(2):708-719.
Yang C, Wang Z, Hollebone BP, Brown CE, Yang Z, Landriault M. 2015. Chromatographic fingerprinting analysis of crude oils and petroleum products. Handbook of oil spill science and technology.93-163.

Zadro S, Haken JK, Pinczewski WV. 1985. Analysis of australian crude oils by high-resolution capillary gas-chromatography mass-spectrometry. Journal of Chromatography. 323(2):305-322.
Zhan DL, Fenn JB. 2000. Electrospray mass spectrometry of fossil fuels. Int J Mass Spectrom. 194(2-3):197-208.
Zheng X, Dupuis KT, Aly NA, Zhou Y, Smith FB, Tang K, Smith RD, Baker ES. 2018. Utilizing ion mobility spectrometry and mass spectrometry for the analysis of polycyclic aromatic hydrocarbons, polychlorinated biphenyls, polybrominated diphenyl ethers and their metabolites. Anal Chim Acta. 1037:265-273.
Zhu GY, Wang M, Chi LX, Li JF, Wu ZH, Zhang ZY. 2020. Discovery and molecular characterization of organic caged compounds and polysulfanes in zhongba81 crude oil, sichuan basin, china. Energ Fuel. 34(6):6811-6821.
Zito P, Podgorski DC, Bartges T, Guillemette F, Roebuck Jr JA, Spencer RGM, Rodgers RP, Tarr MA. 2020. Sunlight-induced molecular progression of oil into oxidized oil soluble species, interfacial material, and dissolved organic matter. Energy Fuels. 34(4):4721-4726.

APPENDIX A

SUPPLEMENTAL FIGURES

Figure S2.1 Gasoline range hydrocarbon data for on-shore crude oil samples analyzed with GC-MS. Sample identifiers are shown above each plot. The bars indicate relative abundance (% abundance) for each feature as compared to the features listed on the x-axis. Abbreviations are: n-C3, propane; i-C4, isobutane; n-C4, butane; i-C5, isopentane; n-C5, pentane; 22DMC4, 2,2-dimethylbutane; CYC5, cyclopentane; 2MC5, 2-methylpentane; 3MC5, 3-methylpentane; n-C6, hexane; MCP, methylcyclopentane; DMP, dimethylcyclopentane; BENZ, benzene; CYC6, cyclohexane; UNK#1, unknown peak; 2MC6, 2-methylhexane; 3MC6, 3-methylhexane; 1CI3 DMCP, 1,cis-3-dimethylcyclopentane; 1TR3 DMCP, 1,trans-3-dimethylcyclopentane; 1TR2 DMCP, 1,trans-2-dimethylcyclopentane; n-C7, hepatane; MCH; methylcyclohexane; TOLU, toluene; n-C8, octane; ETHYLBENZ, ethylbenzene; M-&P-XYL, meta- and para-xylene; O-XYL, ortho-xylene.

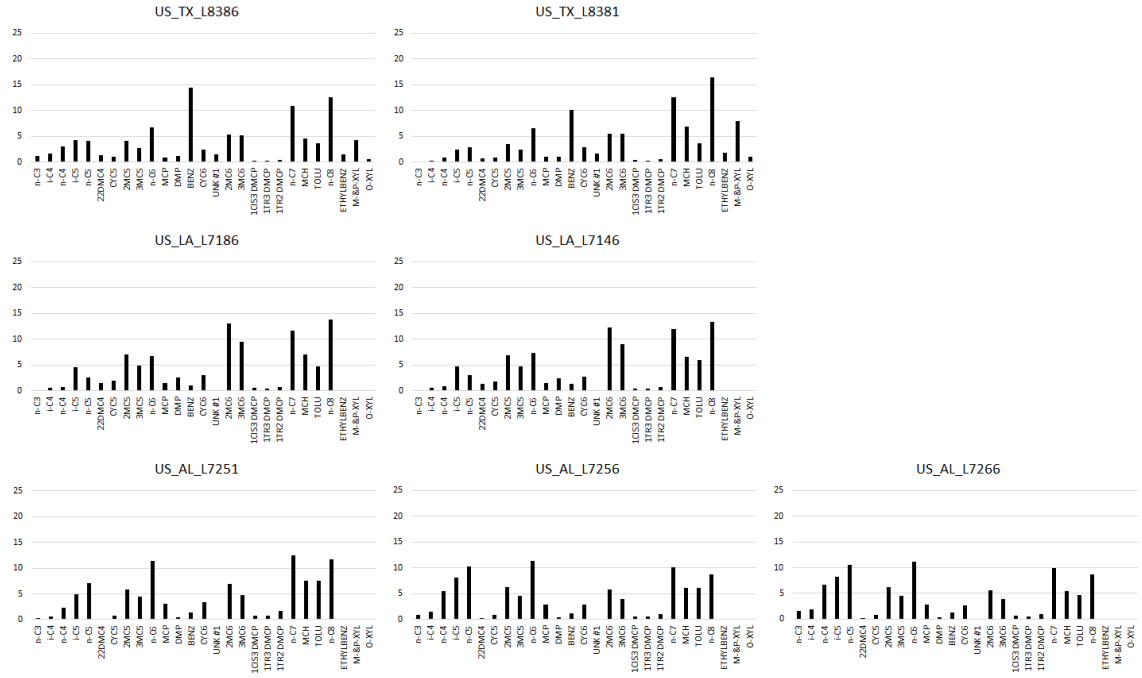


Figure S2.2 Gasoline range hydrocarbon data for off-shore crude oil samples (East Cameron and South Timbalier areas) analyzed with GC-MS. Sample identifiers are shown above each plot. The bars indicate relative abundance (% abundance) for each feature as compared to the features listed on the x-axis. Abbreviations are indicated in the legend to Supplemental Figure 1.

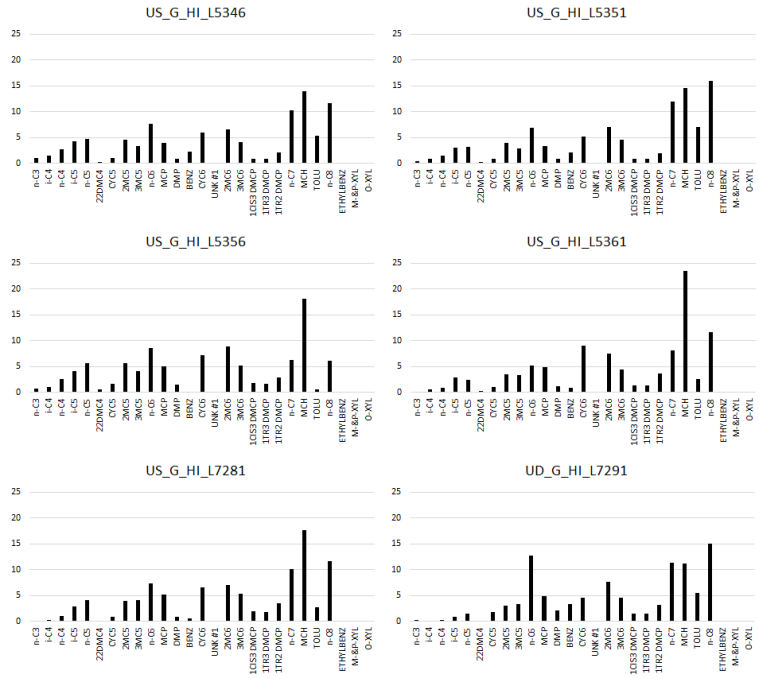


Figure S2.3 Gasoline range hydrocarbon data for off-shore crude oil samples (High Island area) analyzed with GC-MS. Sample identifiers are shown above each plot. The bars indicate relative abundance (% abundance) for each feature as compared to the features listed on the x-axis. Abbreviations are indicated in the legend to Supplemental Figure 1.

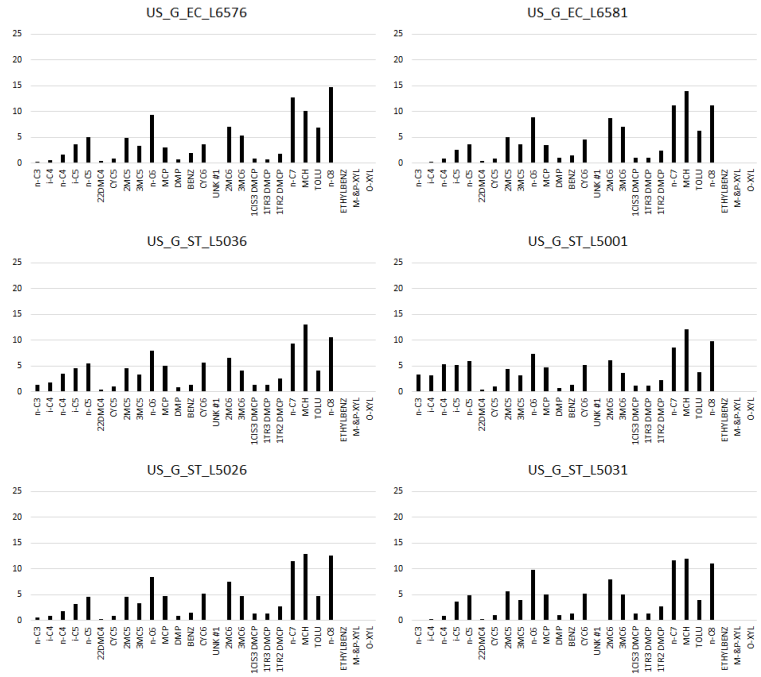


Figure S2.4 Composition of Alkanes from C3-C32 (on-shore samples from TX, LA and AL). The bars indicate relative abundance (% abundance) for each feature as compared to the features listed on the x-axis.

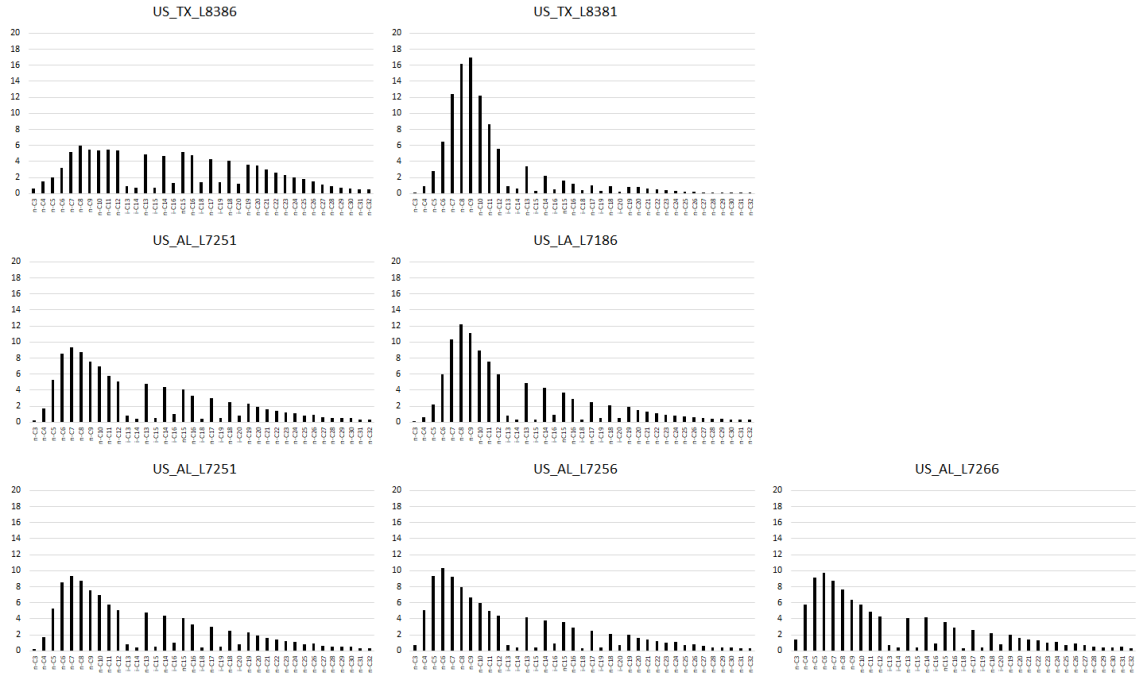


Figure S2.5 Composition of Alkanes from C3-C32 (off-shore samples from HI). The bars indicate relative abundance (% abundance) for each feature as compared to the features listed on the x-axis.

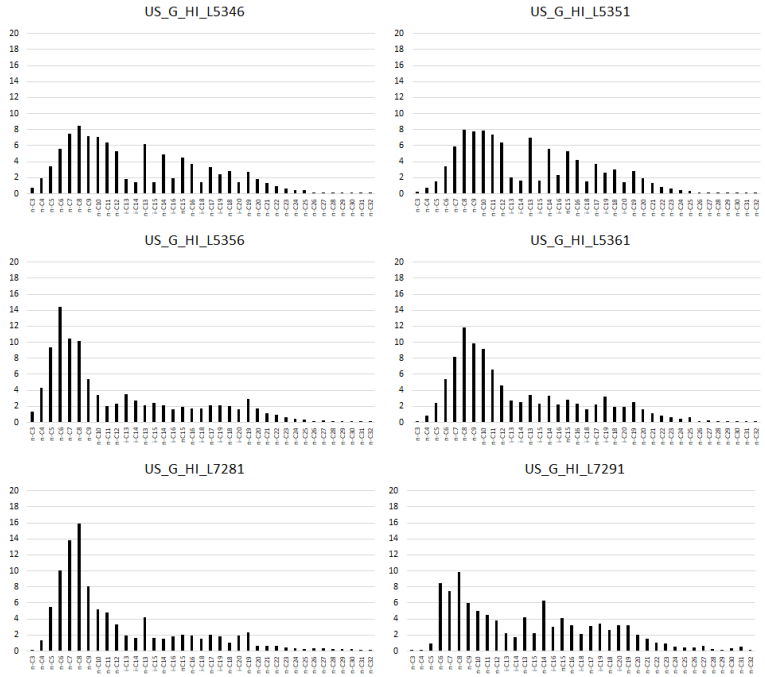


Figure S2.6 Composition of Alkanes from C3-C32 (off-shore samples from EC and ST). The bars indicate relative abundance (% abundance) for each feature as compared to the features listed on the x-axis.

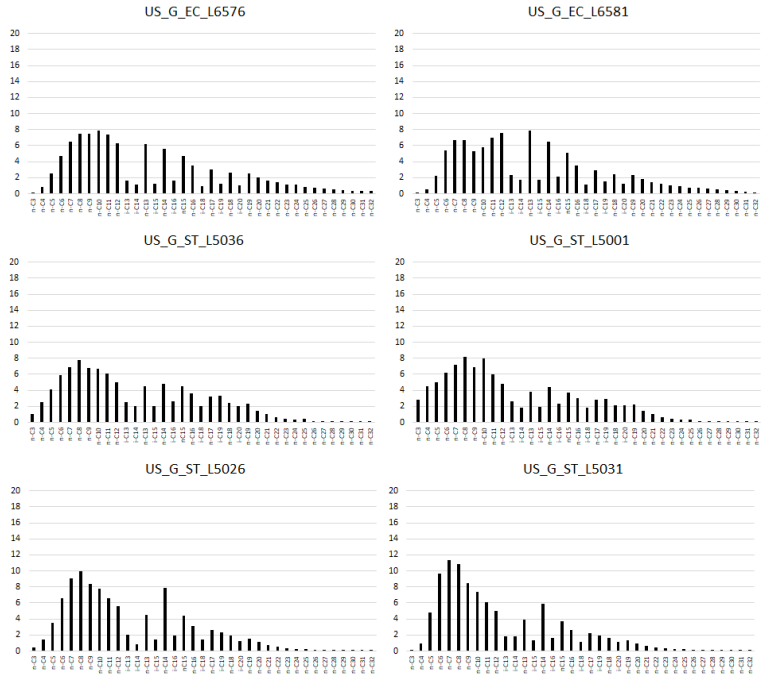


Figure S2.7 GC-MS analysis of n-alkane and isoprenoids (on-shore samples from TX, LA and AL). The bars indicate relative abundance (% abundance) for each feature as compared to the features listed on the x-axis. Abbreviations are: i-C13, isotridecane; i-C14, isotetradecane; i-C15, isopentadecane; i-C16, isohexadecane; i-C18, iso-octadecane.

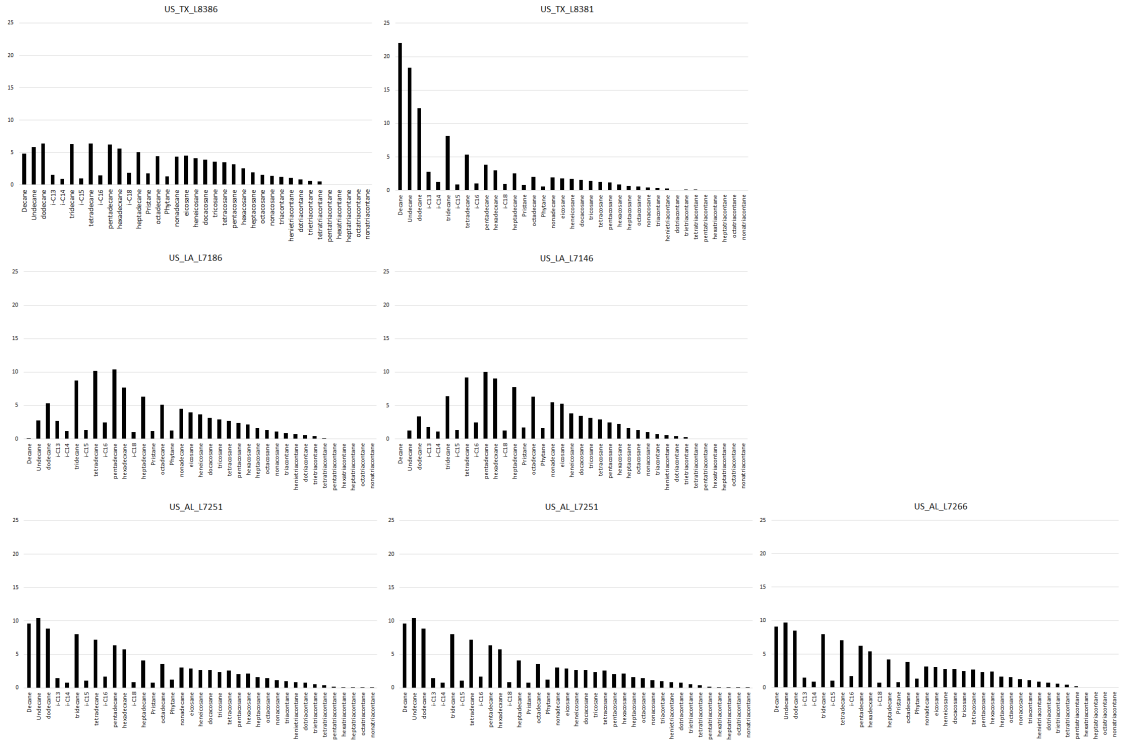


Figure S2.8 GC-MS analysis of n-alkane and isoprenoids (off-shore samples from HI). The bars indicate relative abundance (% abundance) for each feature as compared to the features listed on the x-axis. Abbreviations are indicated in the legend to Supplemental Figure 7.

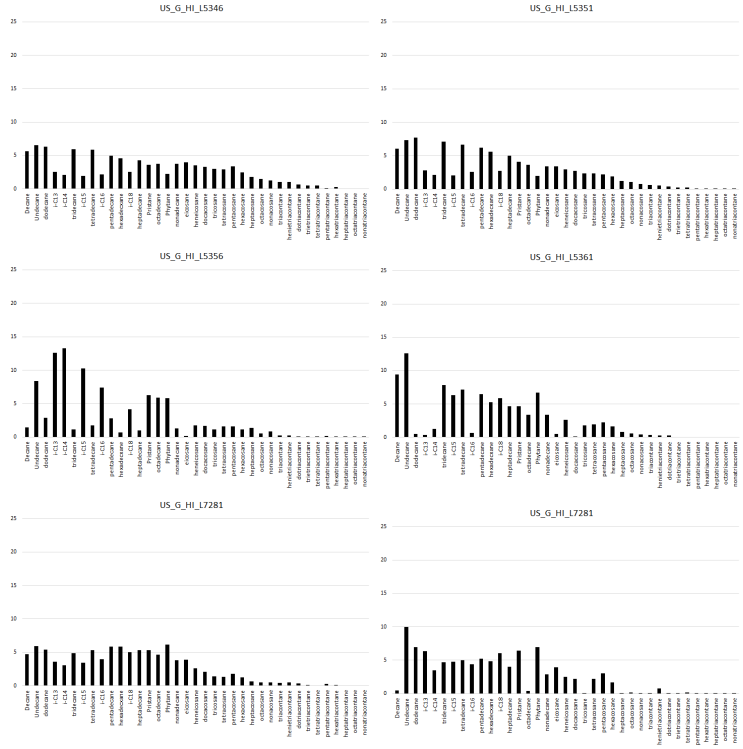


Figure S2.9 GC-MS analysis of n-alkane and isoprenoids (off-shore samples from EC and ST). The bars indicate relative abundance (% abundance) for each feature as compared to the features listed on the x-axis. Abbreviations are indicated in the legend to Supplemental Figure 7.

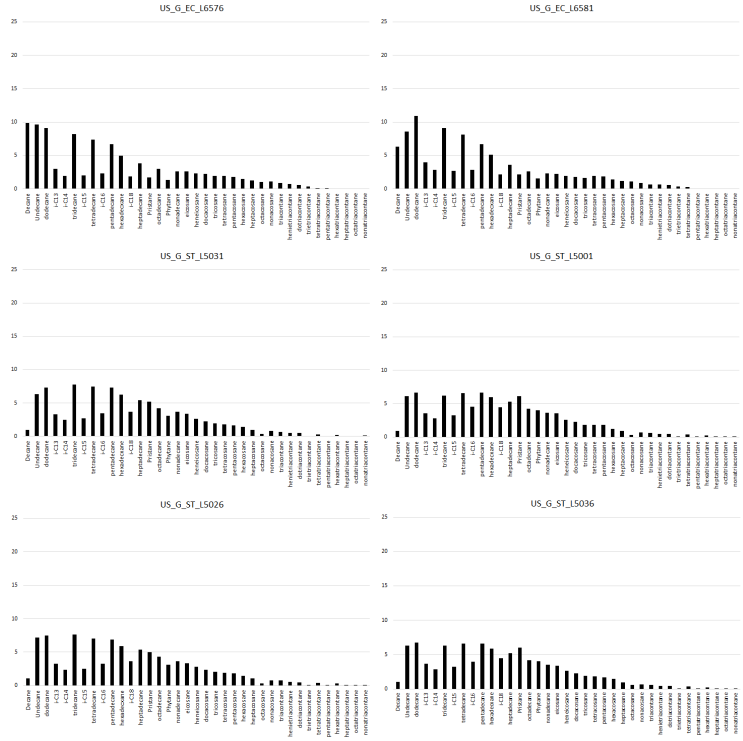


Figure S2.10 GC-MS analysis of PAHs in on-shore samples (TX, LA, AL). The bars indicate relative abundance (% abundance) for each feature as compared to the features listed on the x-axis.

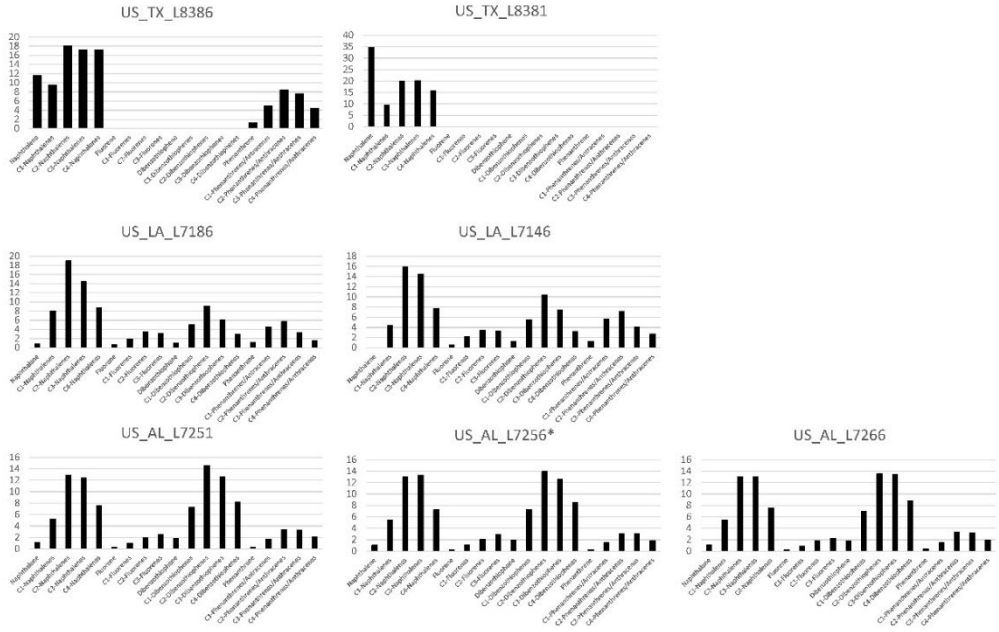


Figure S2.11 GC-MS analysis of PAHs in off-shore samples (EC and ST). The bars indicate relative abundance (% abundance) for each feature as compared to the features listed on the x-axis.

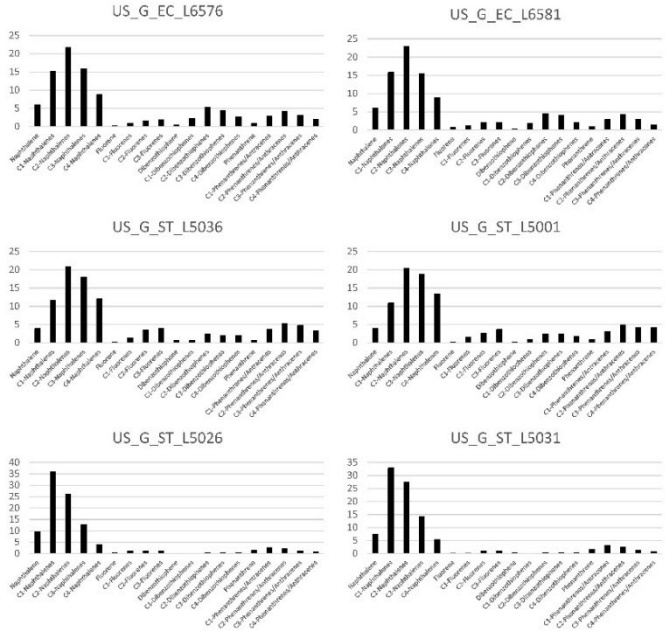


Figure S2.12 GC-MS analysis of PAHs in off-shore samples (HI). The bars indicate relative abundance (% abundance) for each feature as compared to the features listed on the x-axis.

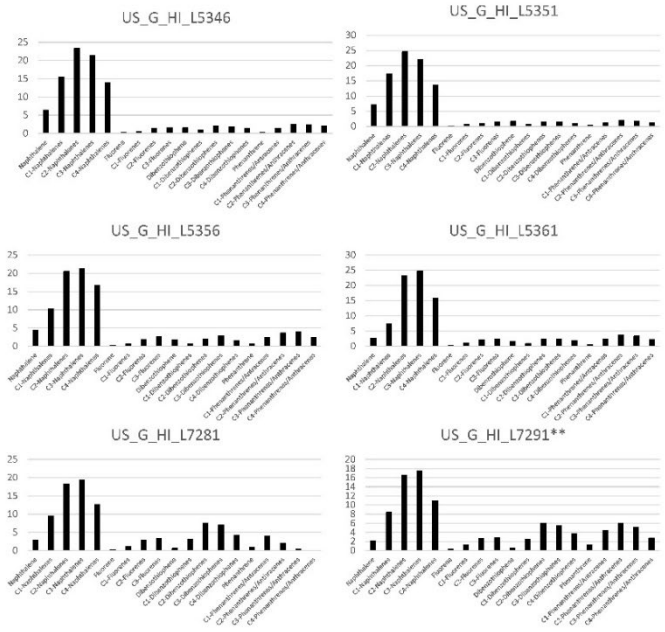


Figure S2.13 Euclidean cluster dendrogram of all samples with GC-MS n-alkane and isoprenoids dataset (Table S1.11).

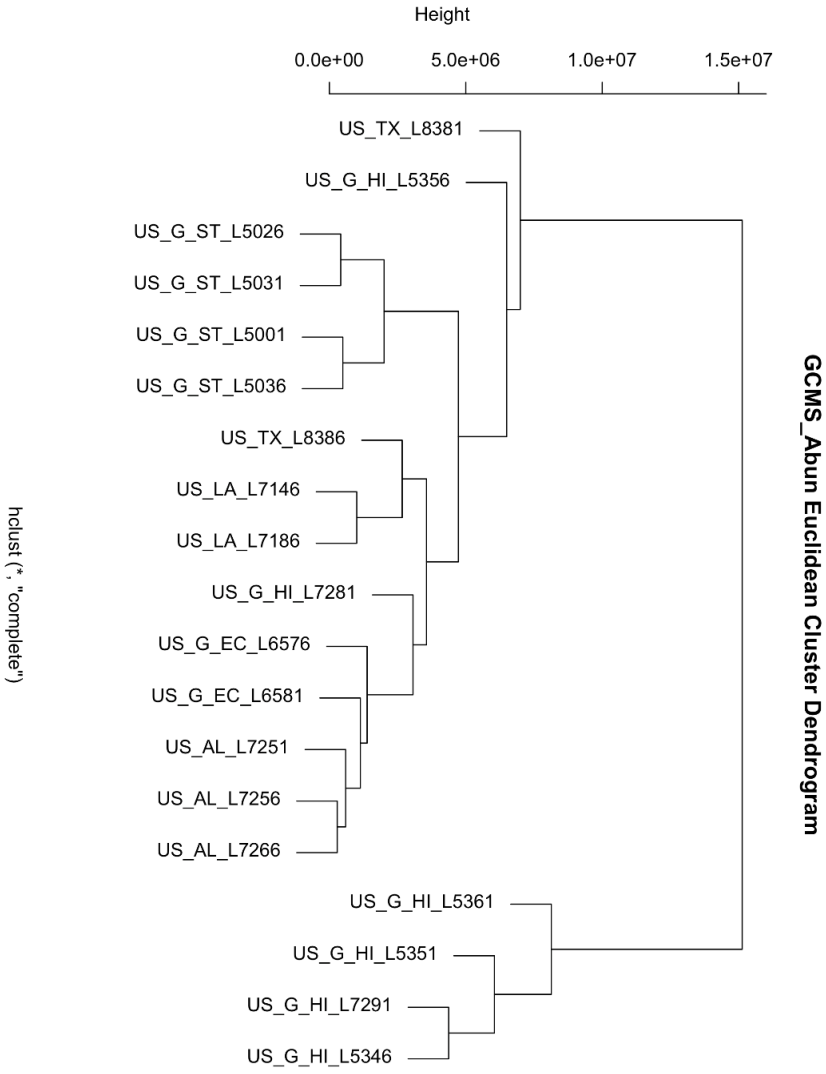


Figure S2.14 Euclidean cluster dendrogram of all samples with GC-MS n-alkane and isoprenoids dataset log-transformed (Table S1.11).

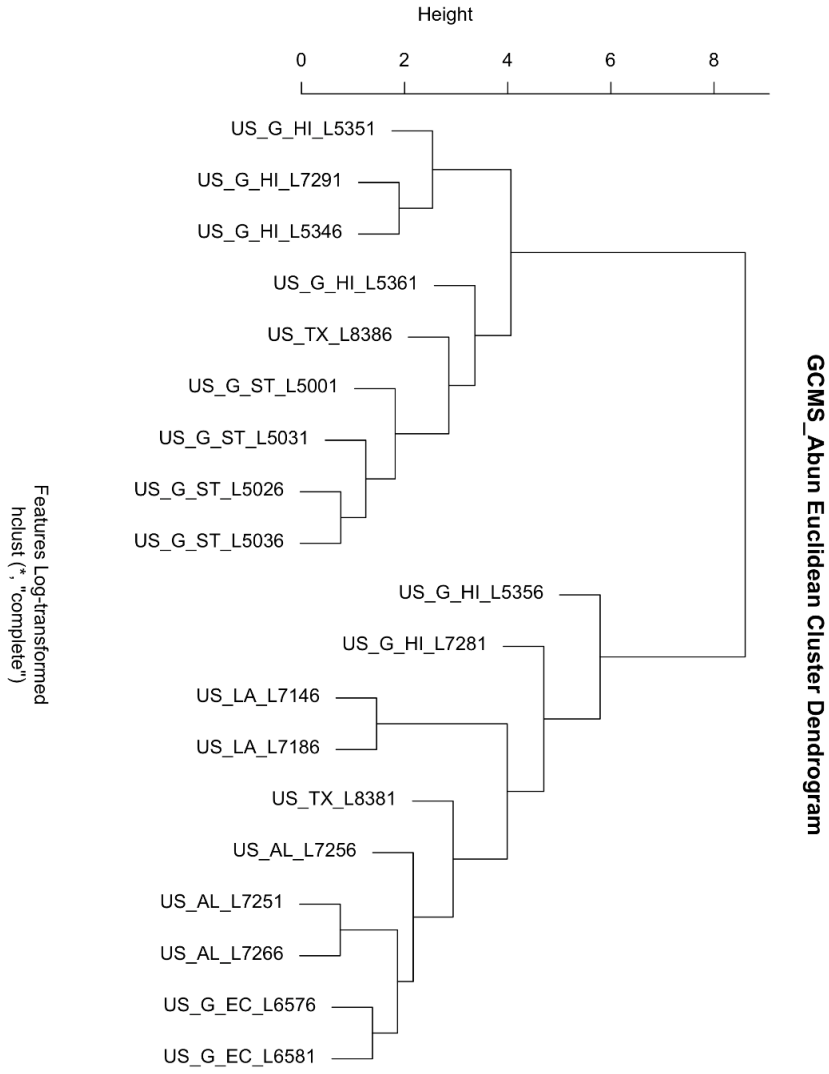


Figure S2.15 Euclidean cluster dendrogram of all samples with GC-MS n-alkane and isoprenoids dataset fractionated abundance (Table S1.11).

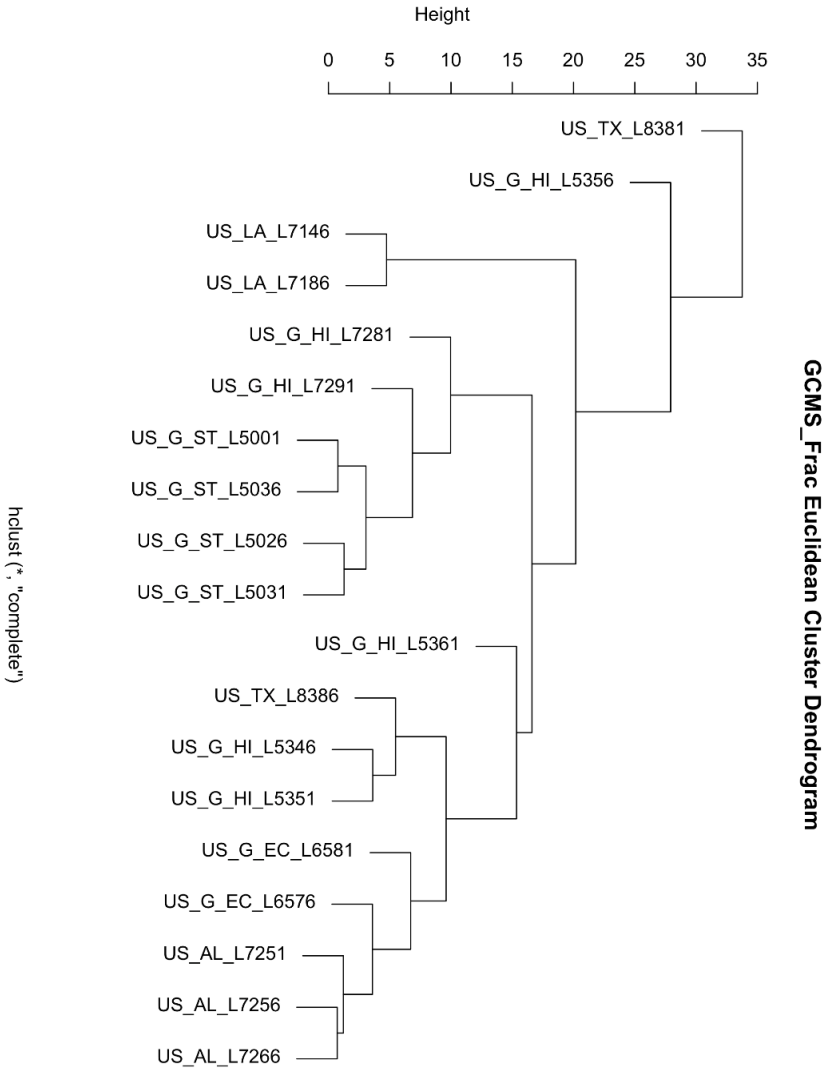


Figure S2.16 Euclidean cluster dendrogram of all samples with GC-MS n-alkane and isoprenoids dataset fractionated abundance log-transformed (Table S1.11).

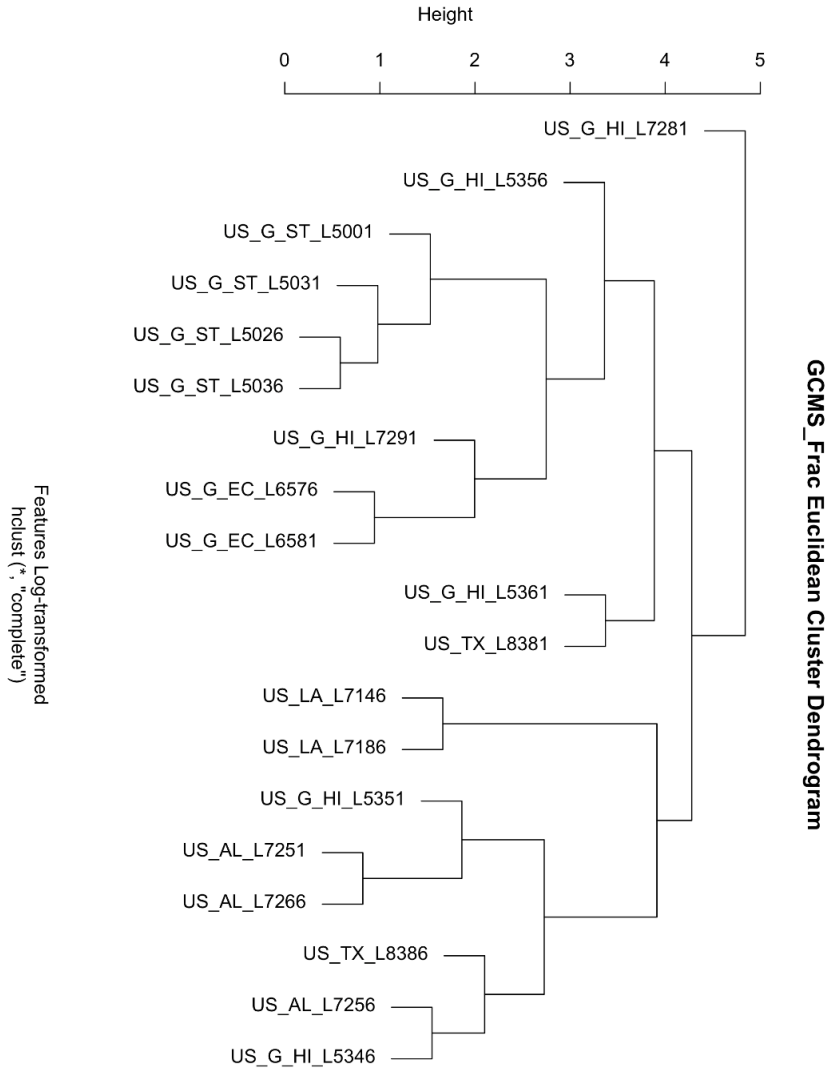


Figure S2.17 Euclidean cluster dendrogram of all samples with GC-MS PAHs dataset (Table S1.12).

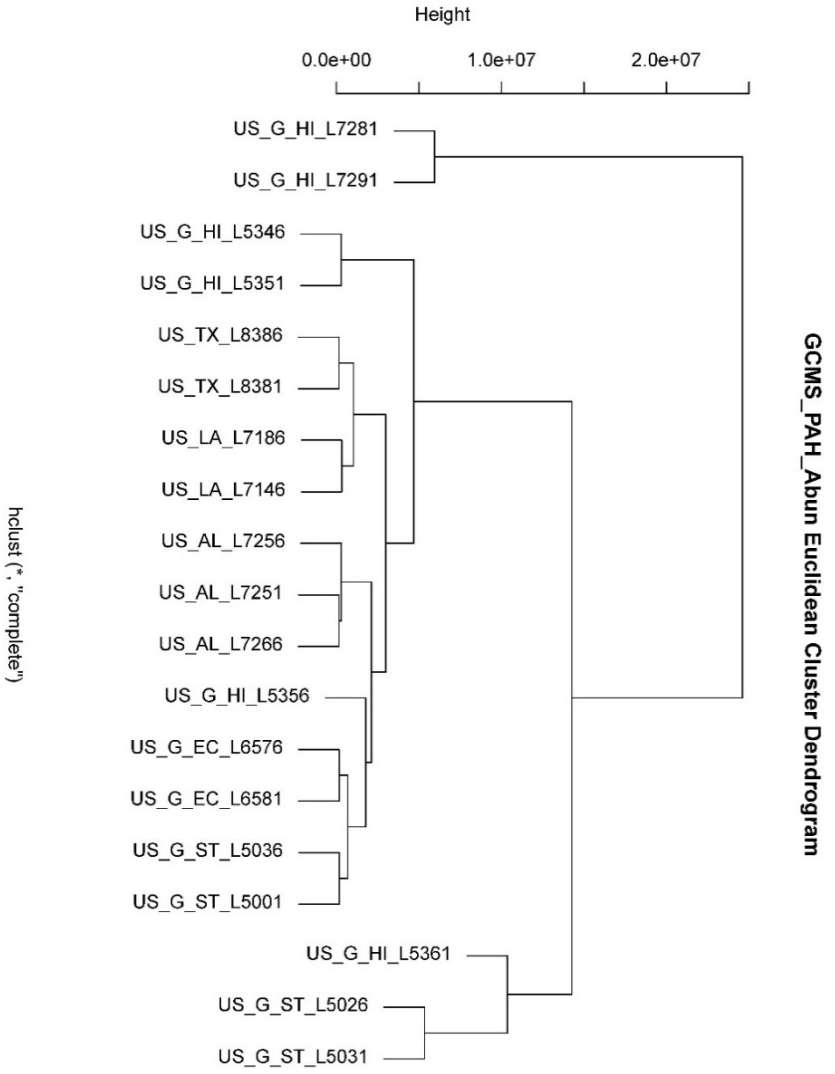


Figure S2.18 Euclidean cluster dendrogram of all samples with GC-MS PAHs dataset (Table S1.12).

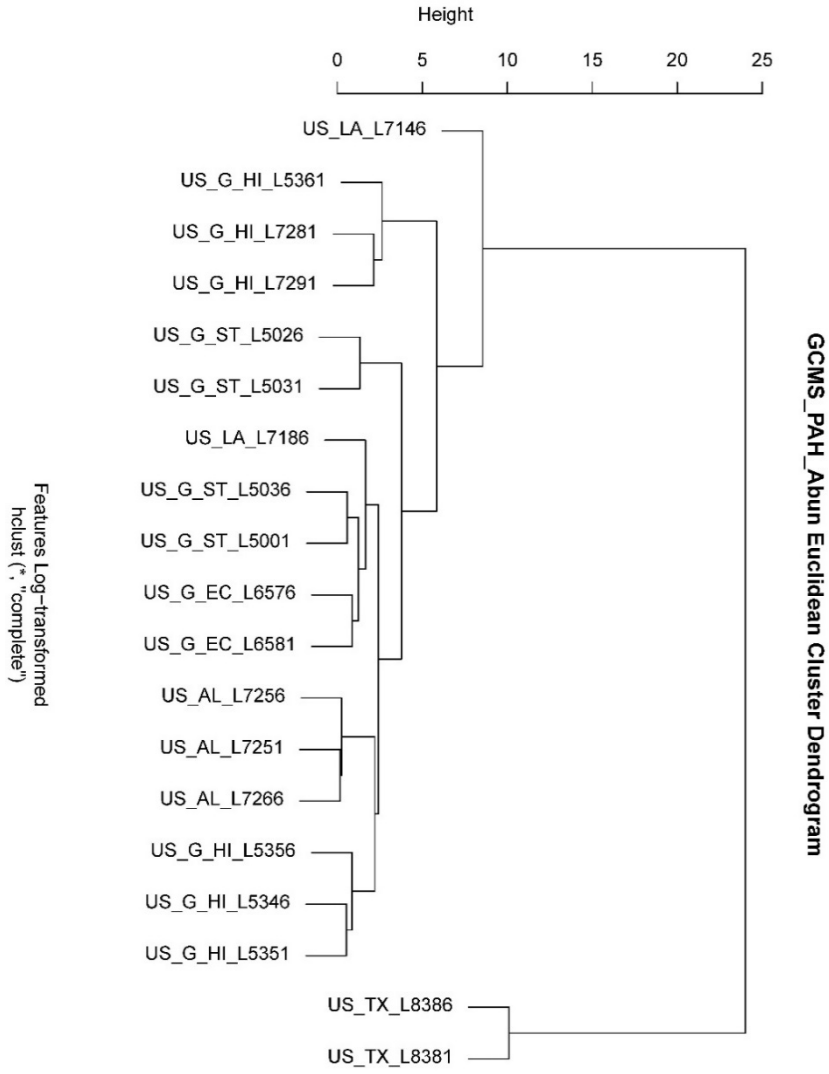


Figure S2.19 Euclidean cluster dendrogram of all samples with GC-MS PAHs dataset (Table S1.12).

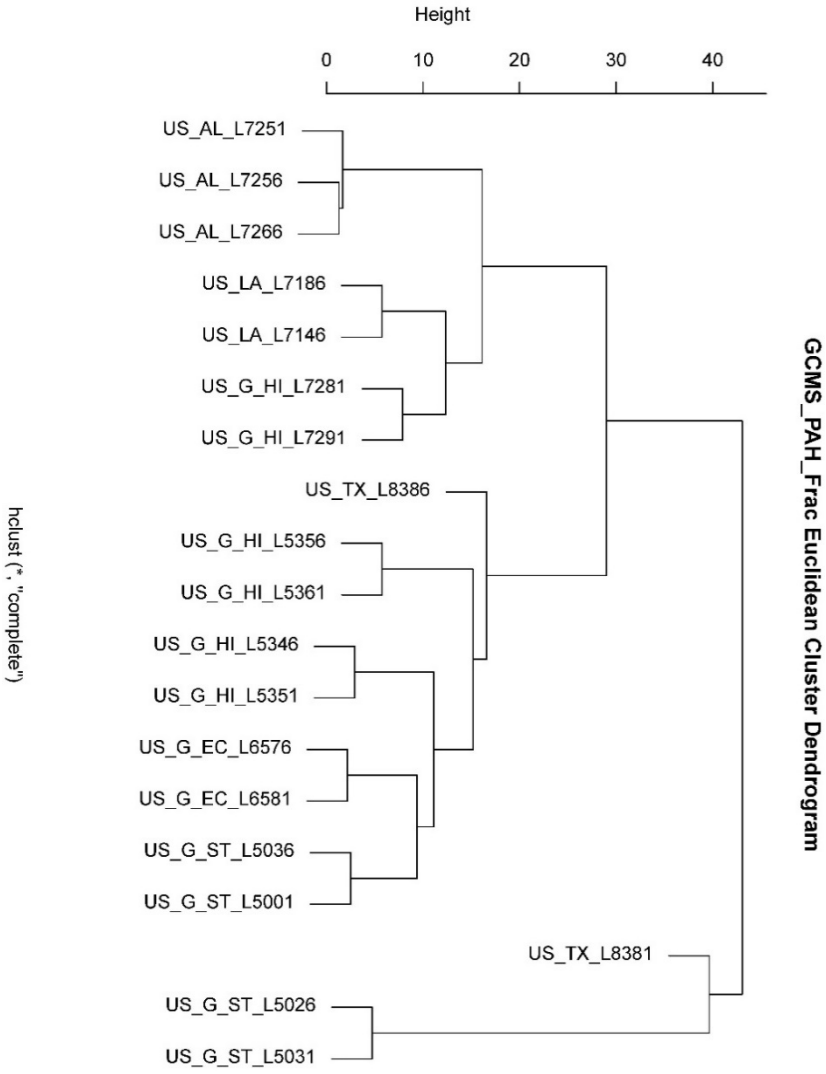


Figure S2.20 Euclidean cluster dendrogram of all samples with GC-MS PAHs dataset (Table S1.12).

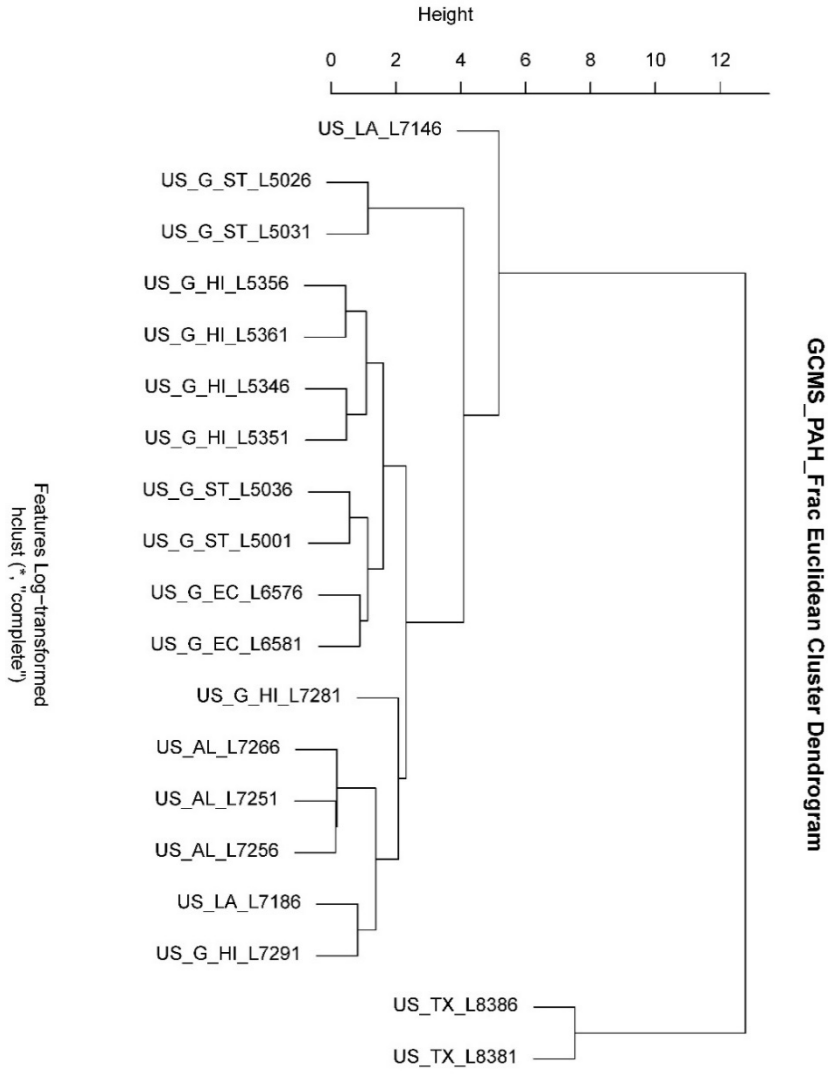


Figure S2.21 Euclidean cluster dendrogram of all samples with all biomarker datasets (Table S1.3-8).

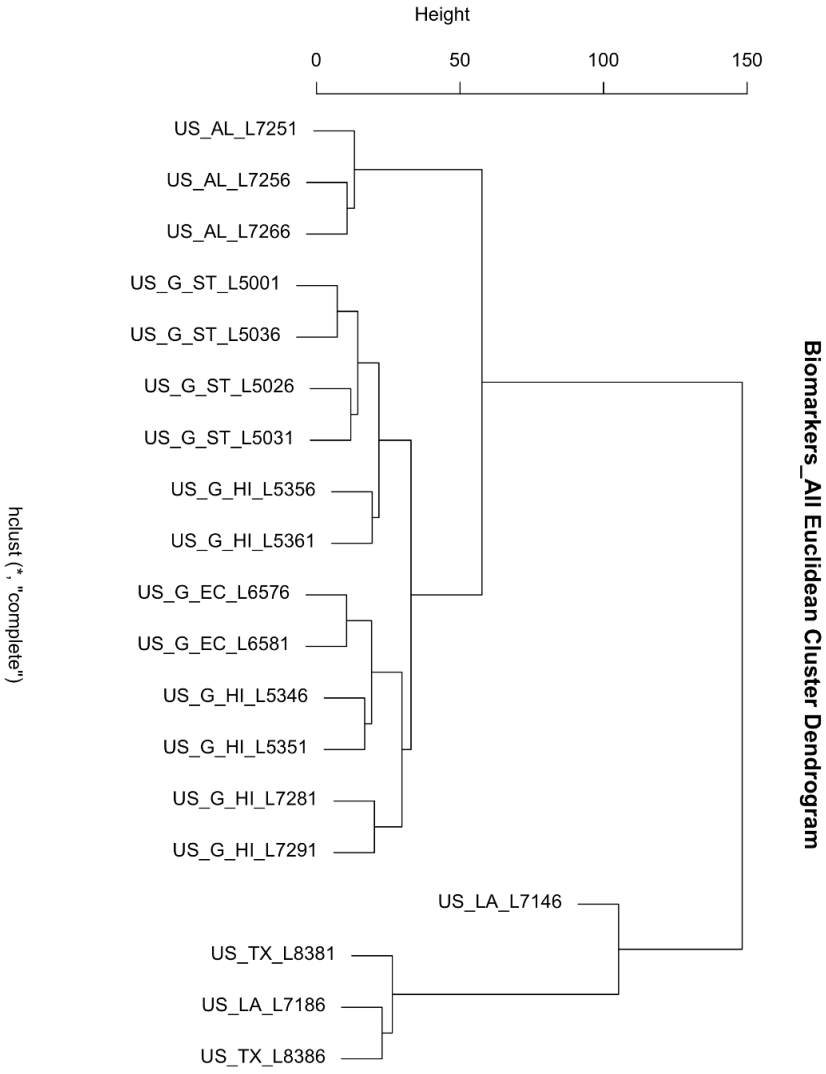


Figure S2.22 Euclidean cluster dendrogram of all samples with all biomarker datasets log-transformed (Table S1.3-8).

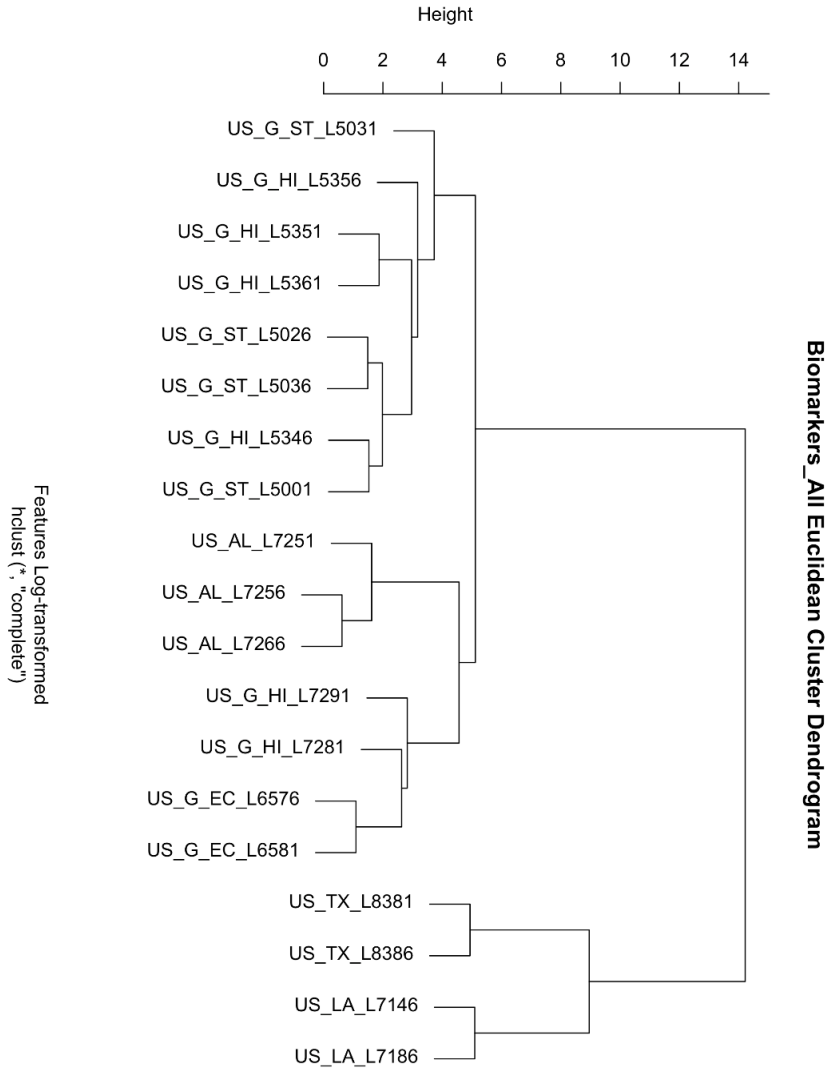


Figure S2.23 Euclidean cluster dendrogram of all samples with Gasoline Range Hydrocarbon dataset (Table S1.9).

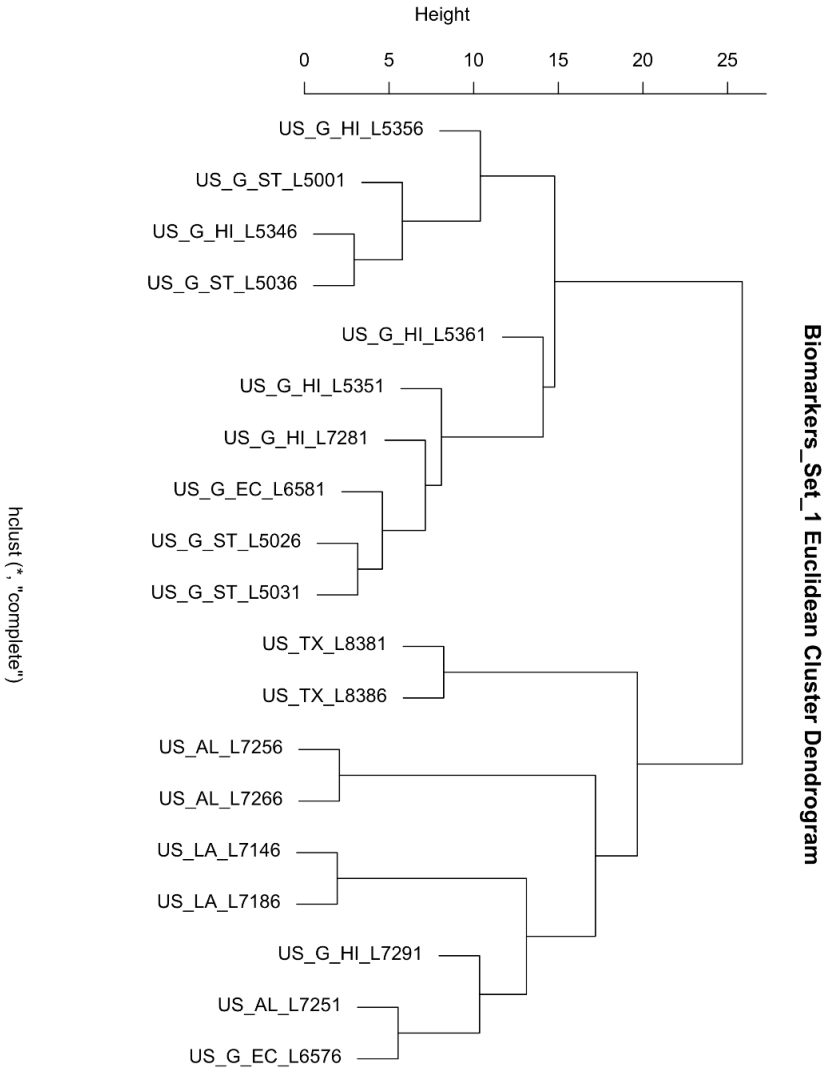


Figure S2.24 Euclidean cluster dendrogram of all samples with Gasoline Range Hydrocarbon dataset log-transformed (Table S1.9).

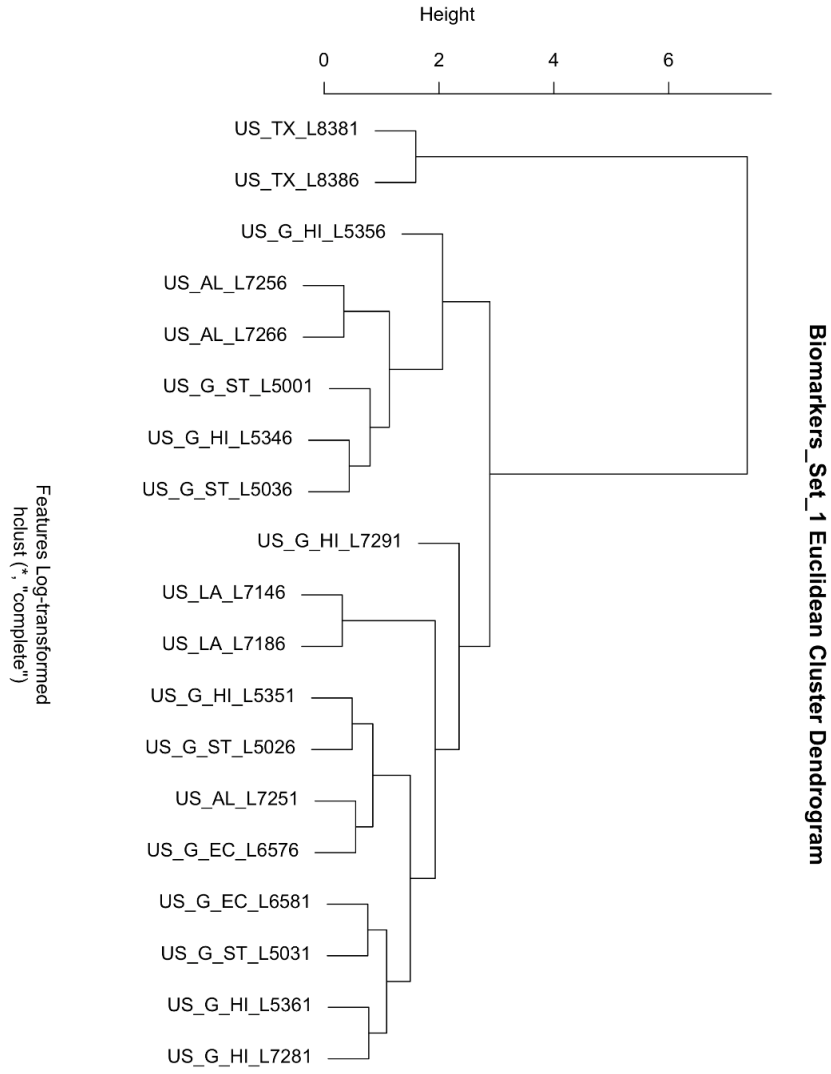


Figure S2.25 Euclidean cluster dendrogram of all samples with %Composition Alkanes C3-C32 dataset (Table S1.10).

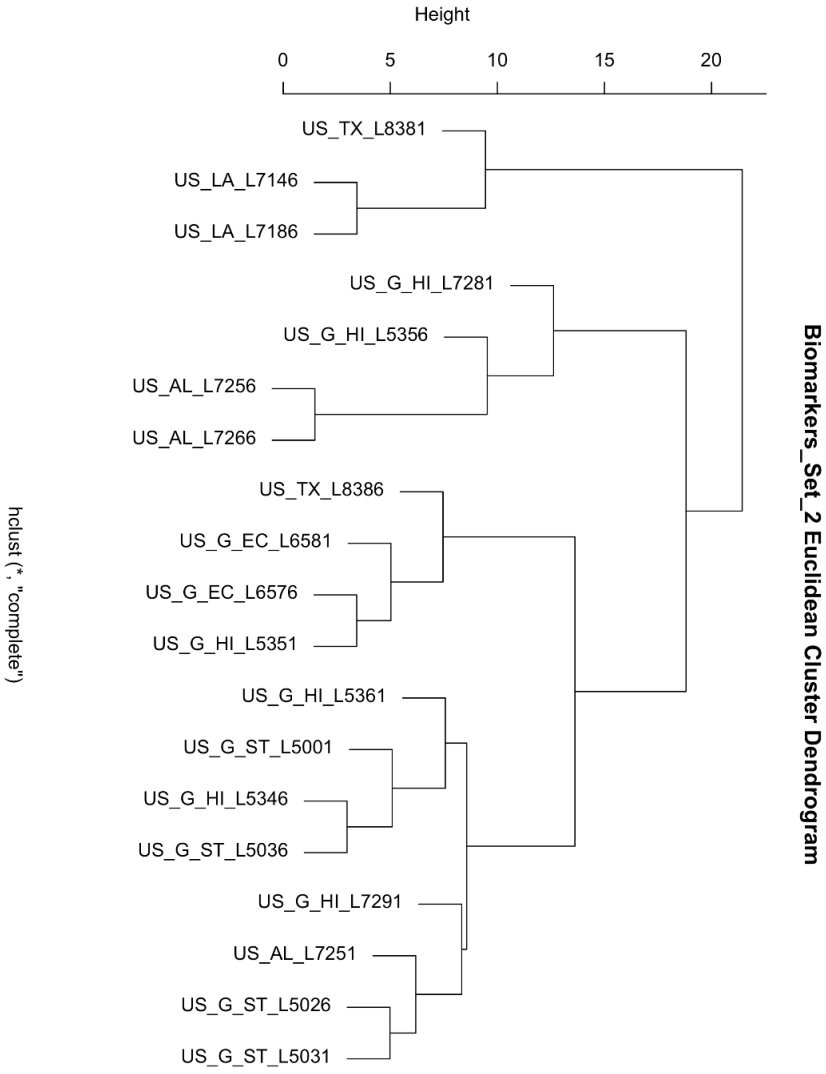


Figure S2.26 Euclidean cluster dendrogram of all samples with %Composition Alkanes C3-C32 dataset log-transformed (Table S1.10).

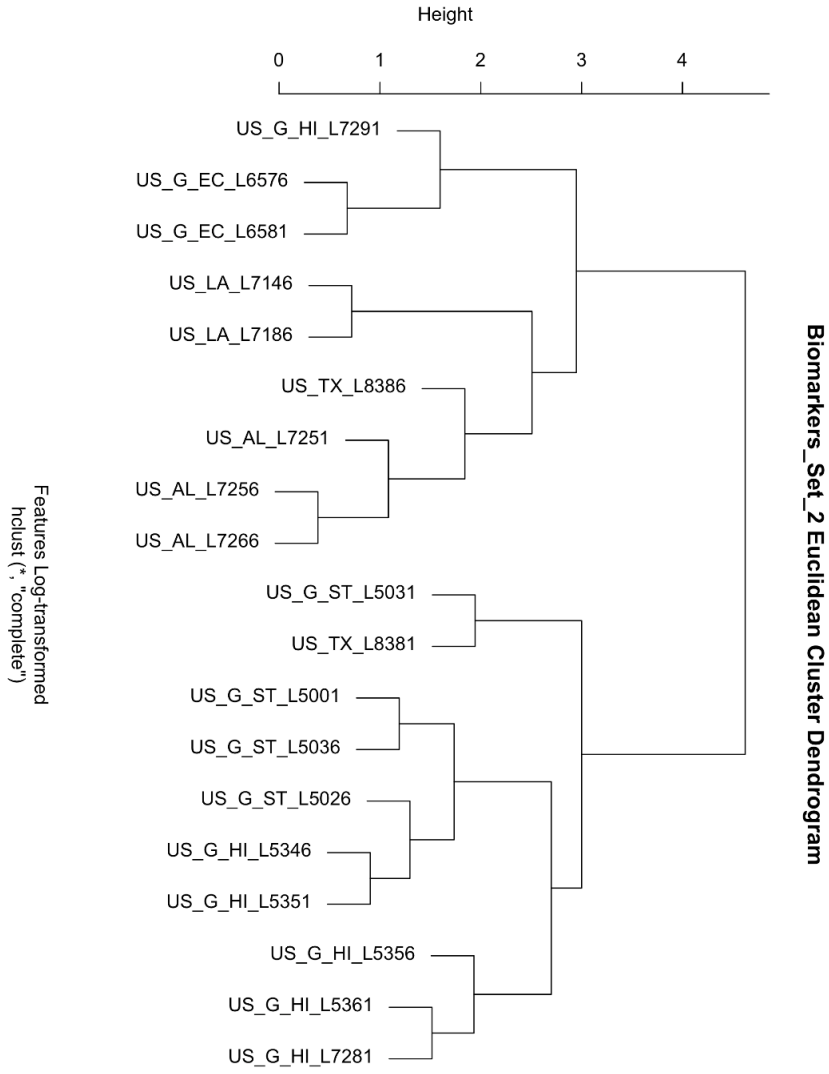


Figure S2.27 Euclidean cluster dendrogram of all samples with all GC-MS data generated measurements (n-alkanes, biomarkers, etc.) dataset log-transformed (Table S1.3-11).

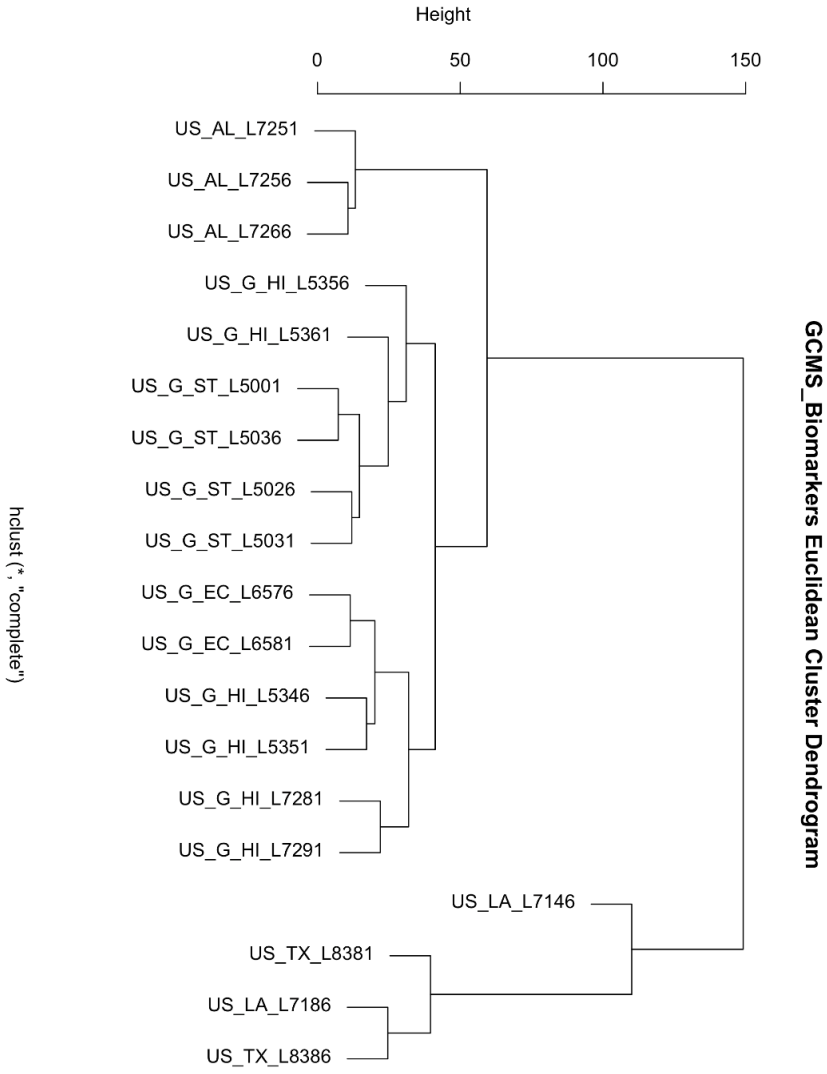


Figure S2.28 Euclidean cluster dendrogram of all samples with all GC-MS data generated measurements (n-alkanes, biomarkers, etc.) datasets log-transformed (Table S1.3-11).

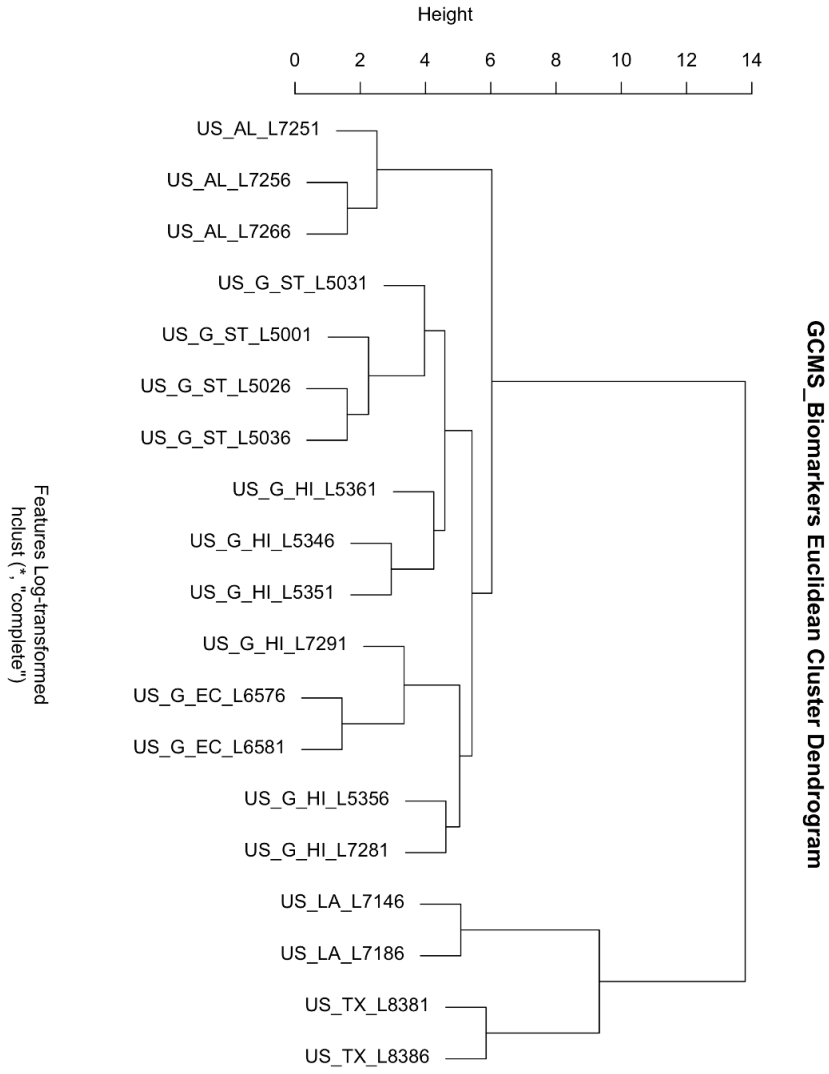


Figure S2.29 Binary cluster of samples in replicates with all features detected through IMS-MS (Table S1.13).

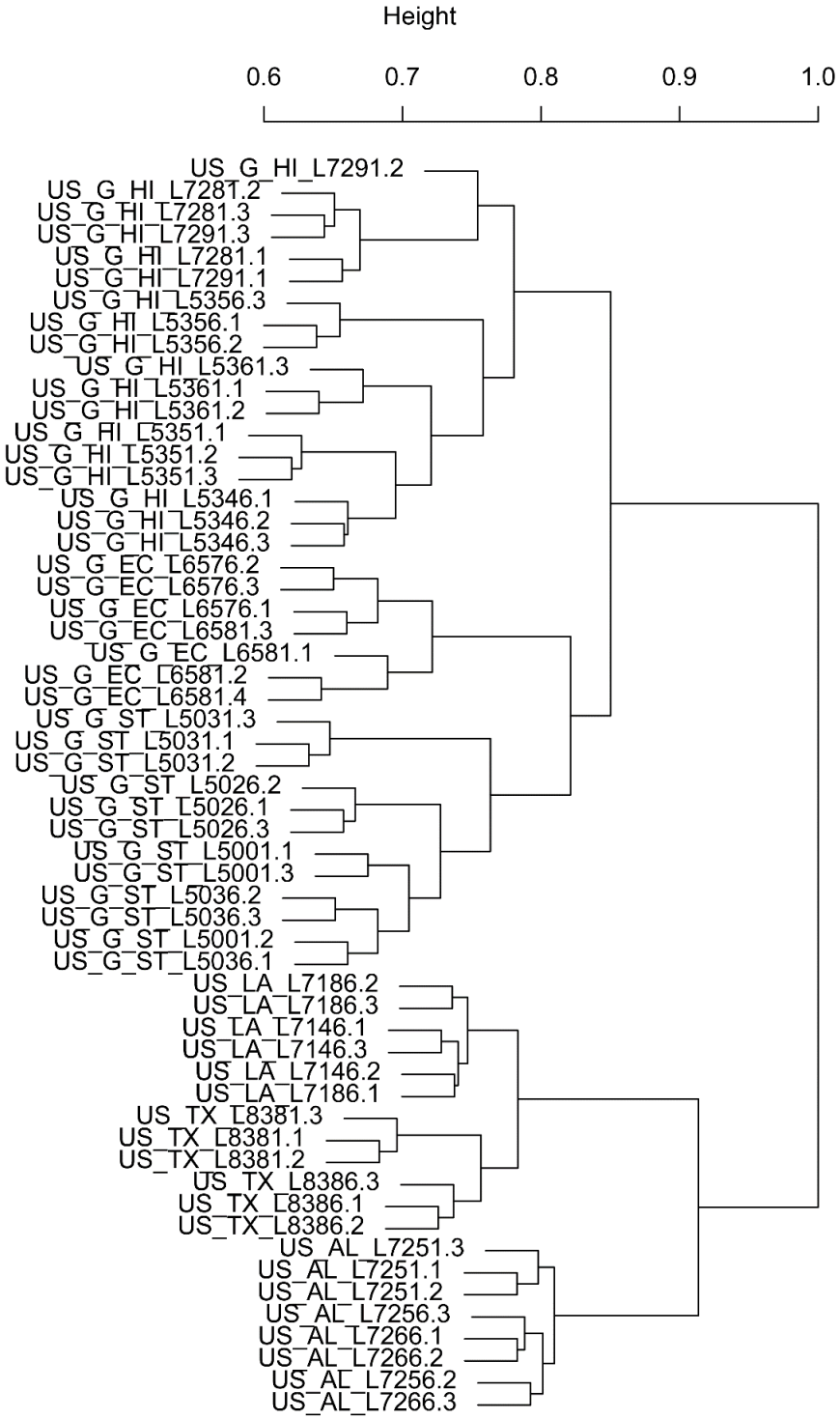


Figure S2.30 Binary cluster of samples in replicates with all features detected through IMS-MS log-transformed abundance (Table S1.13).

Figure S2.31 Binary cluster of samples in replicates with all features detected through IMS-MS normalized log-transformed abundance (Table S1.13).

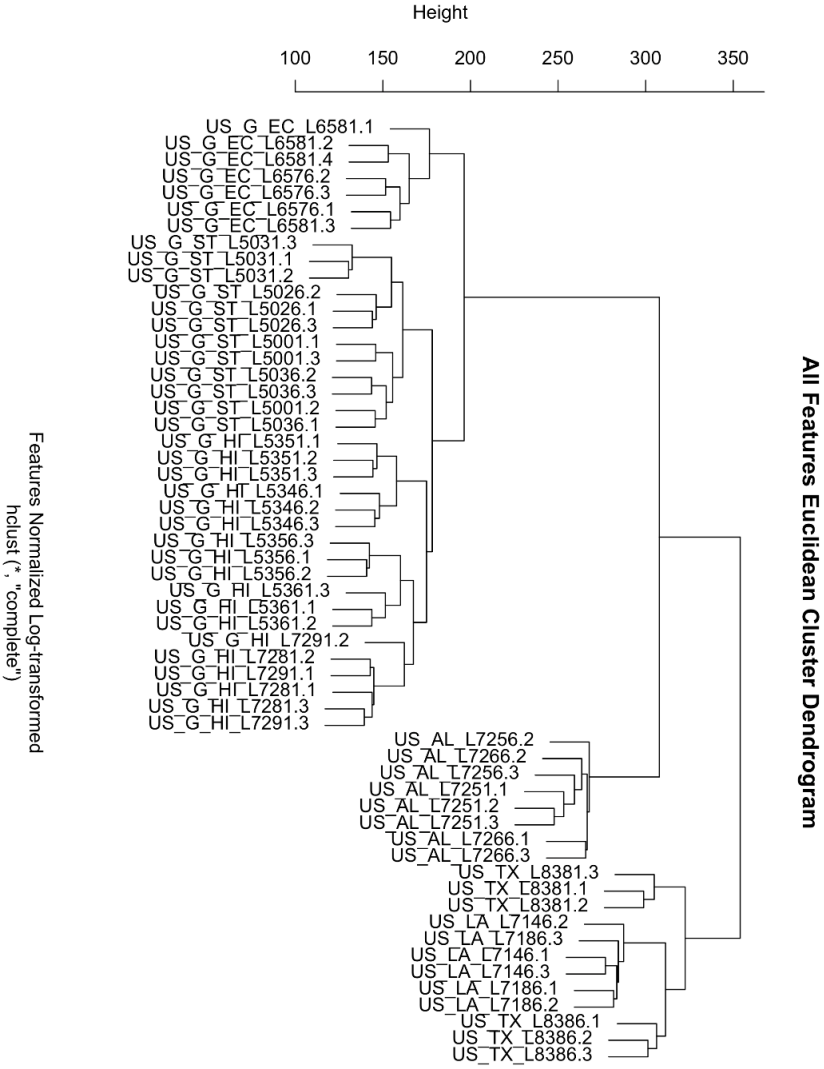


Figure S2.32 Binary cluster of samples with features detected through IMS-MS targeted dataset log-transformed abundance (Table S1.14).

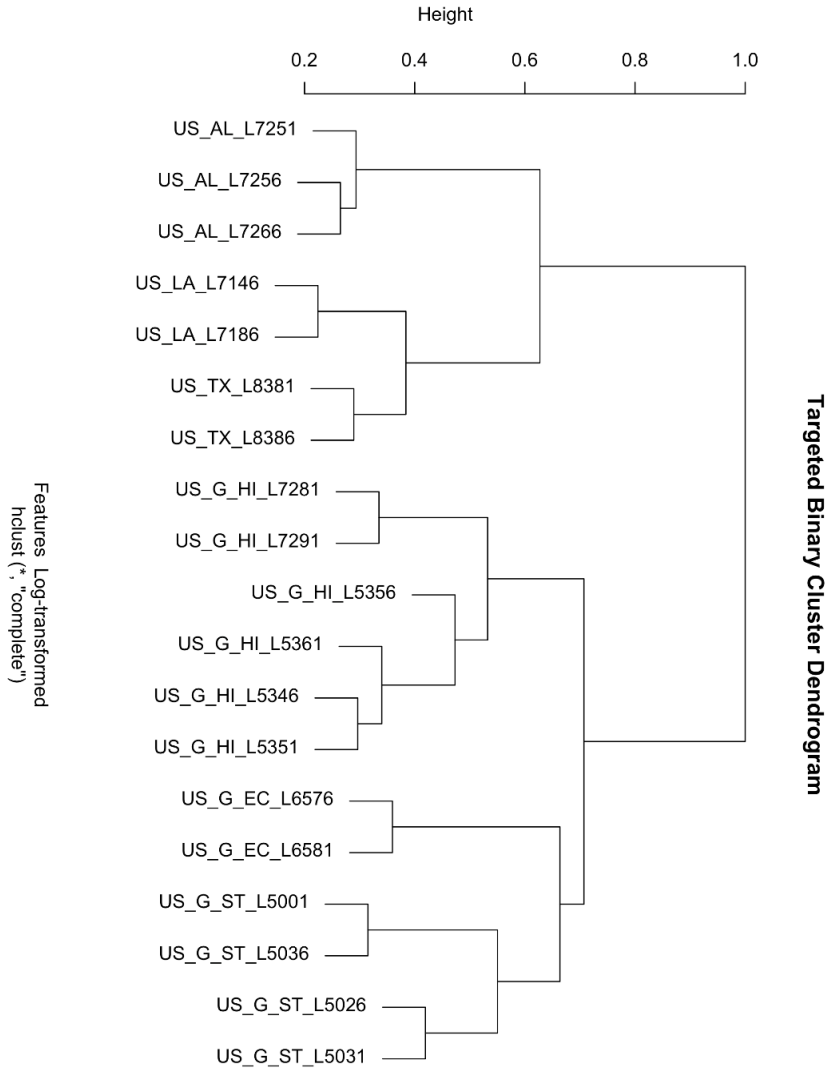


Figure S2.33 Euclidean cluster of samples with features detected through IMS-MS targeted dataset log-transformed abundance (Table S1.14).

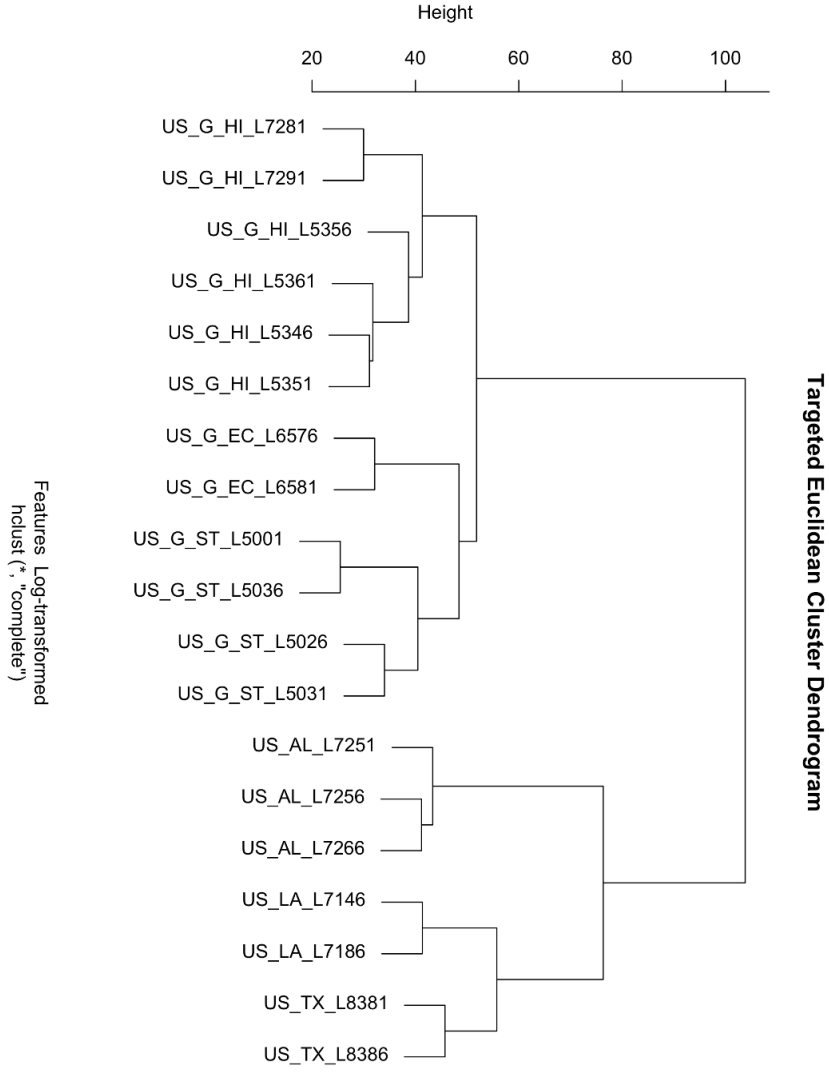


Figure S2.34 Euclidean cluster of samples with features detected through IMS-MS targeted dataset normalized log-transformed abundance (Table S1.14).

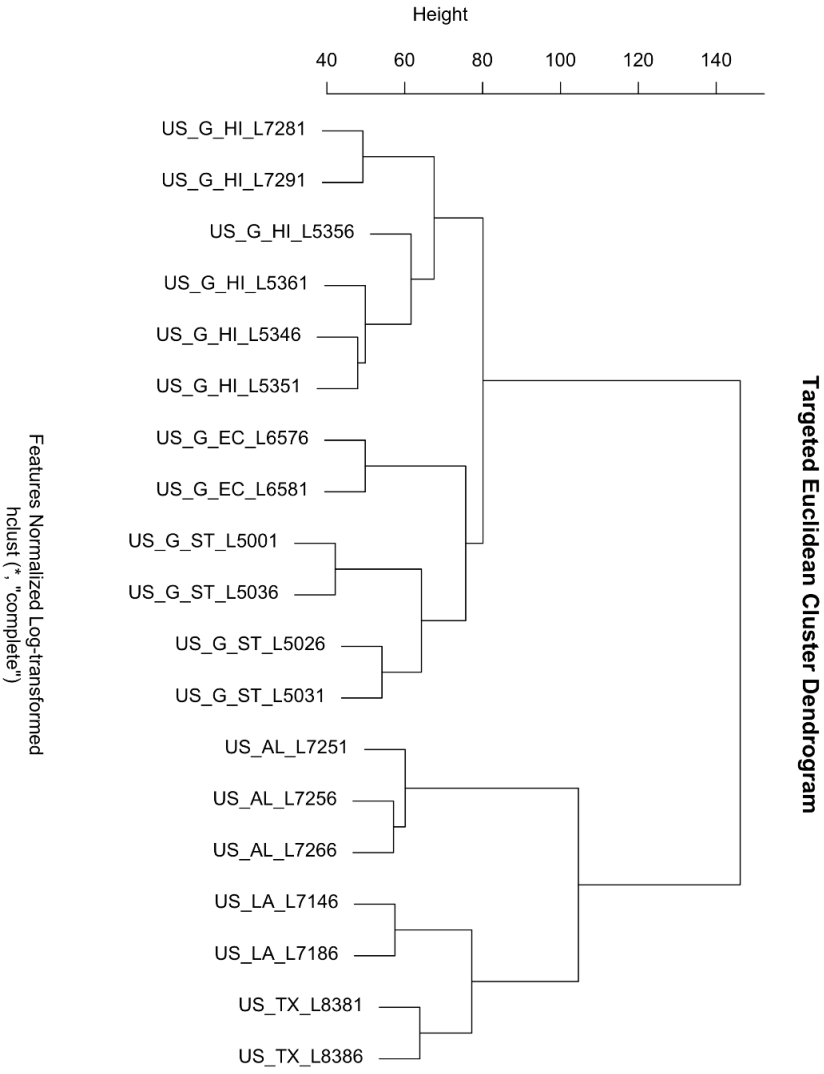


Figure S2.35 Binary cluster of samples with features detected through IMS-MS untargeted dataset log-transformed abundance (Table S1.14).

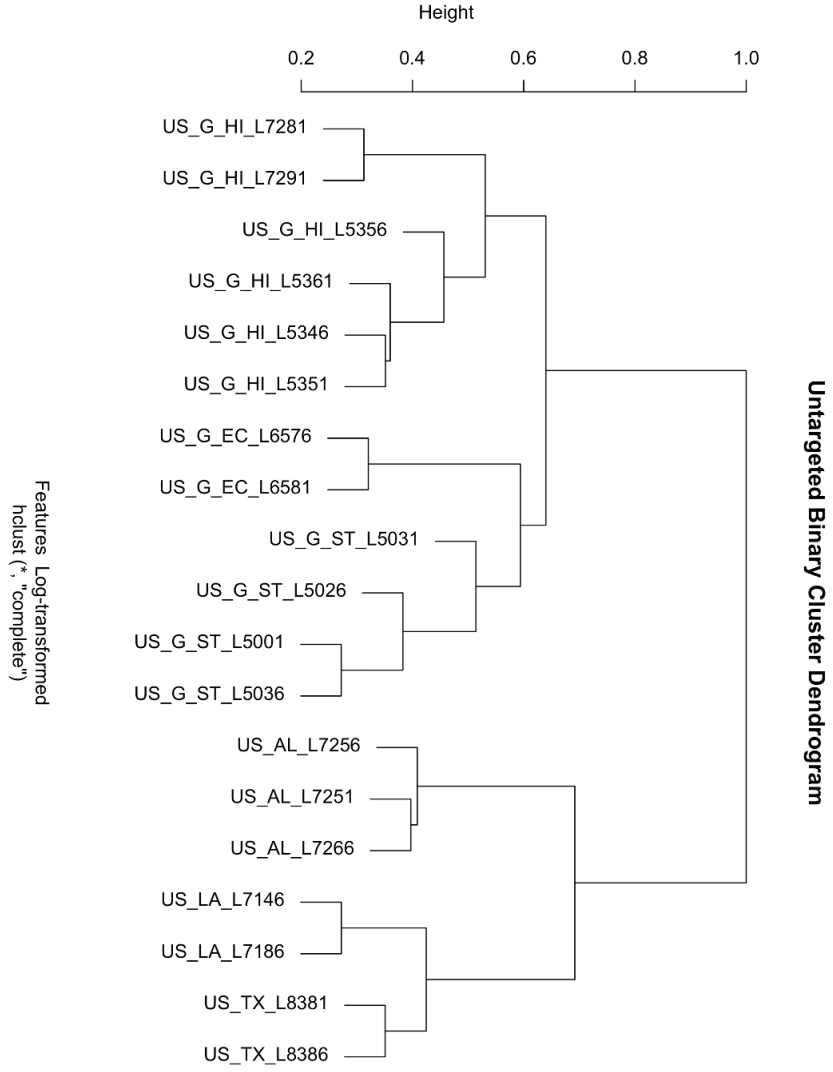


Figure S2.36 Euclidean cluster of samples with features detected through IMS-MS untargeted dataset log-transformed abundance (Table S1.14).

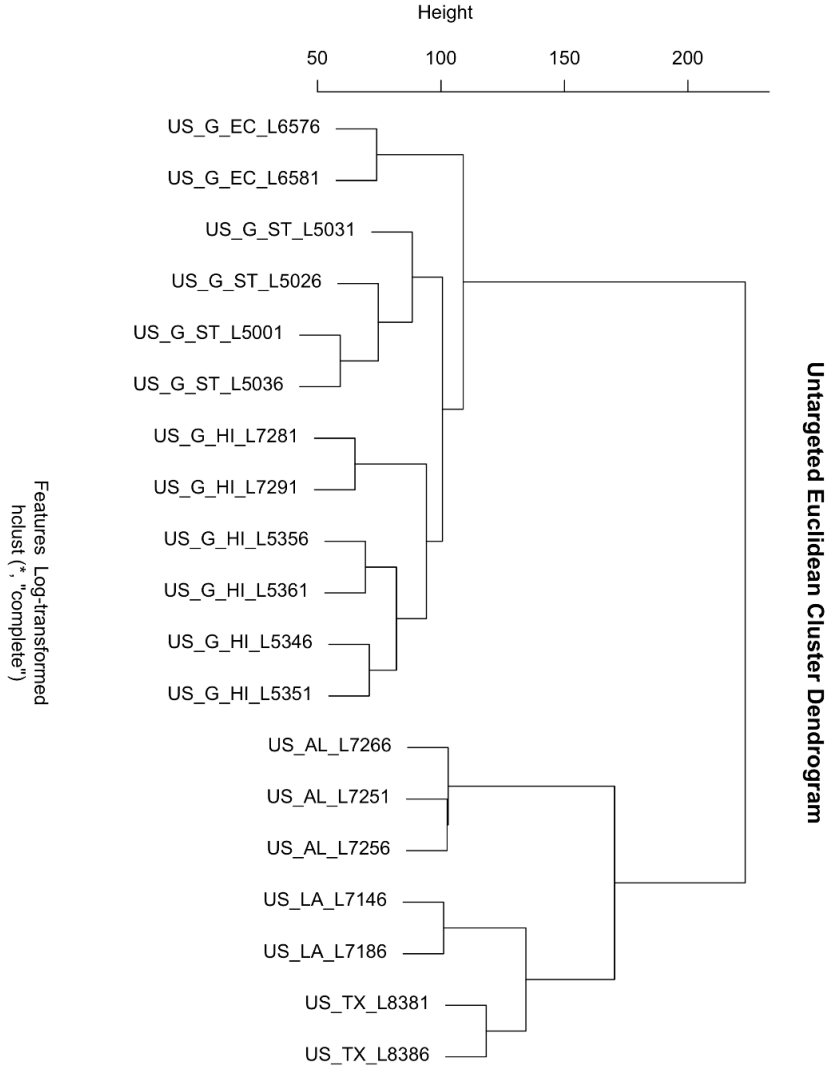


Figure S2.37 Euclidean cluster of samples with features detected through IMS-MS untargeted dataset normalized log-transformed abundance (Table S1.14).

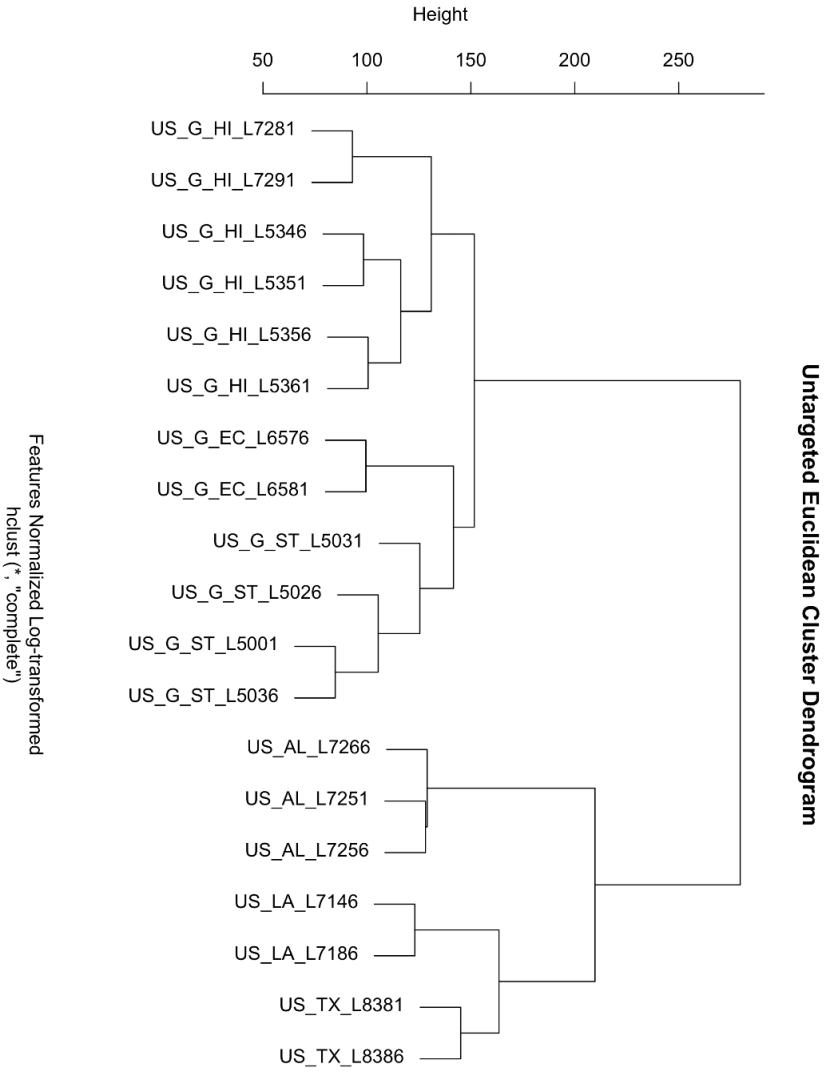


Figure S3.1 GC-MS full scan TIC for all of the petroleum substances evaluated.

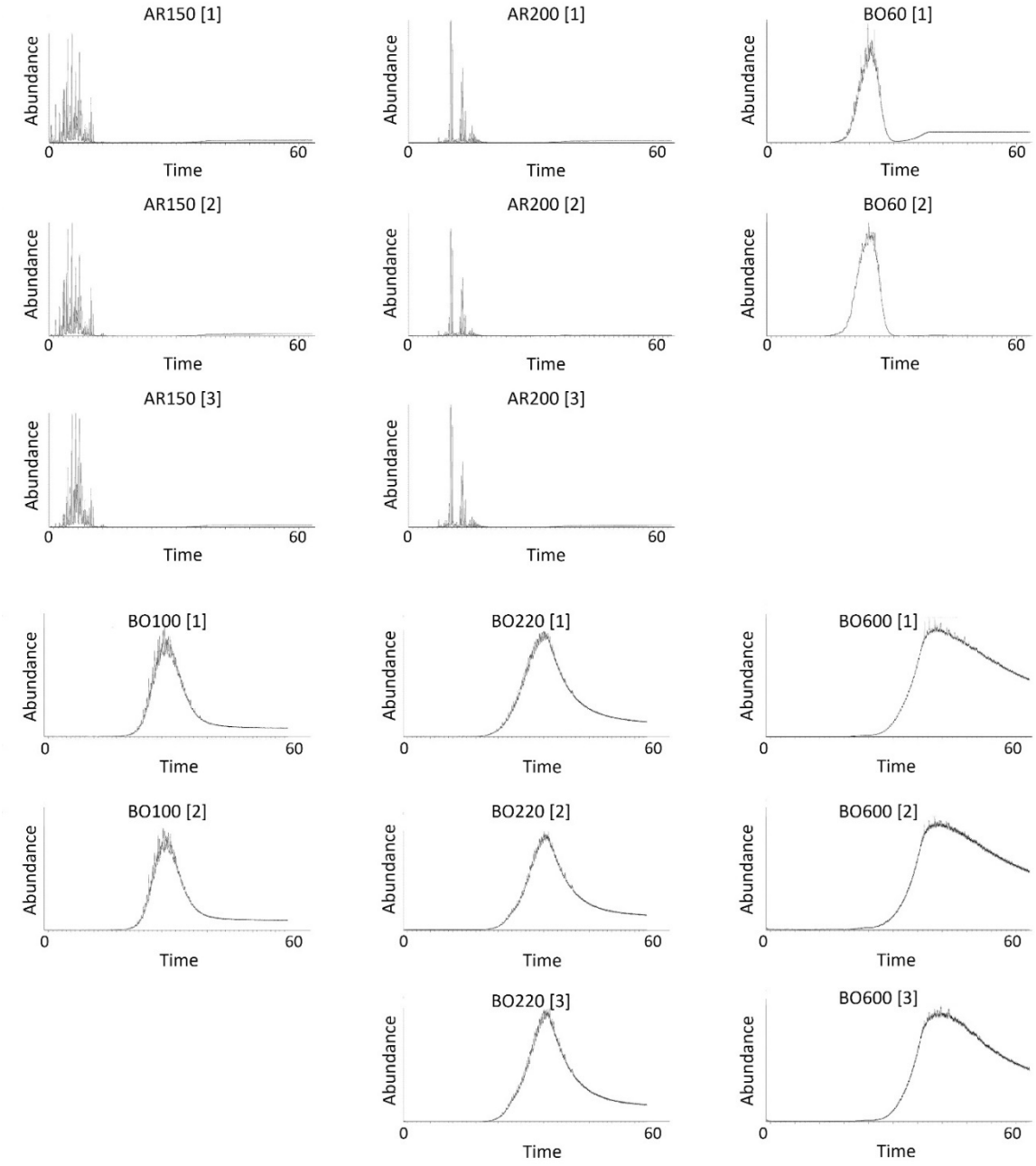


Figure S3.2 IMS-MS spectra for all of the petroleum substances evaluated.

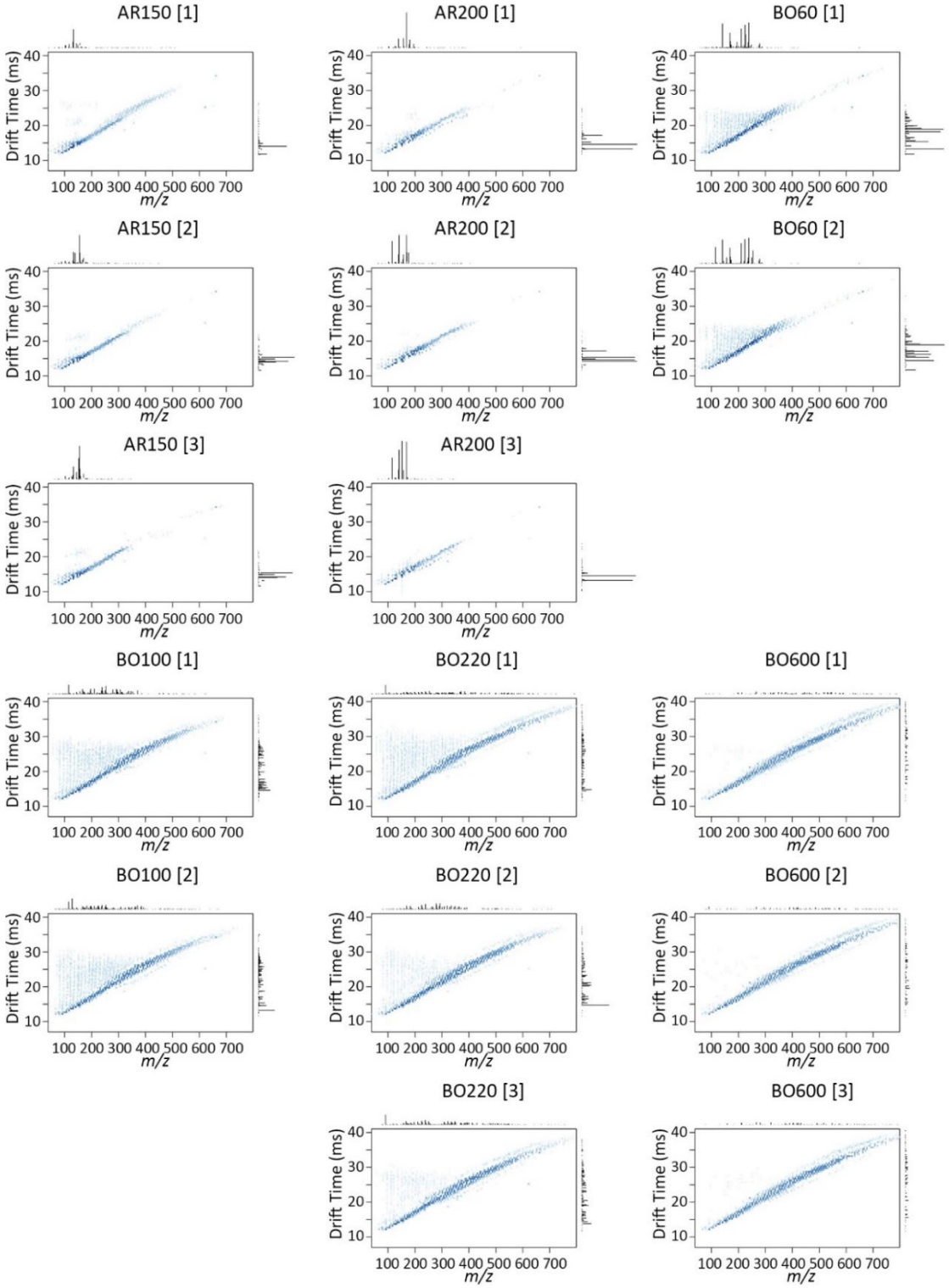


Figure S3.3 Hierarchical clustering dendogram of the petroleum refining products with their technical triplicates based on their chemical profile acquired through IMS-MS-APPI(+).

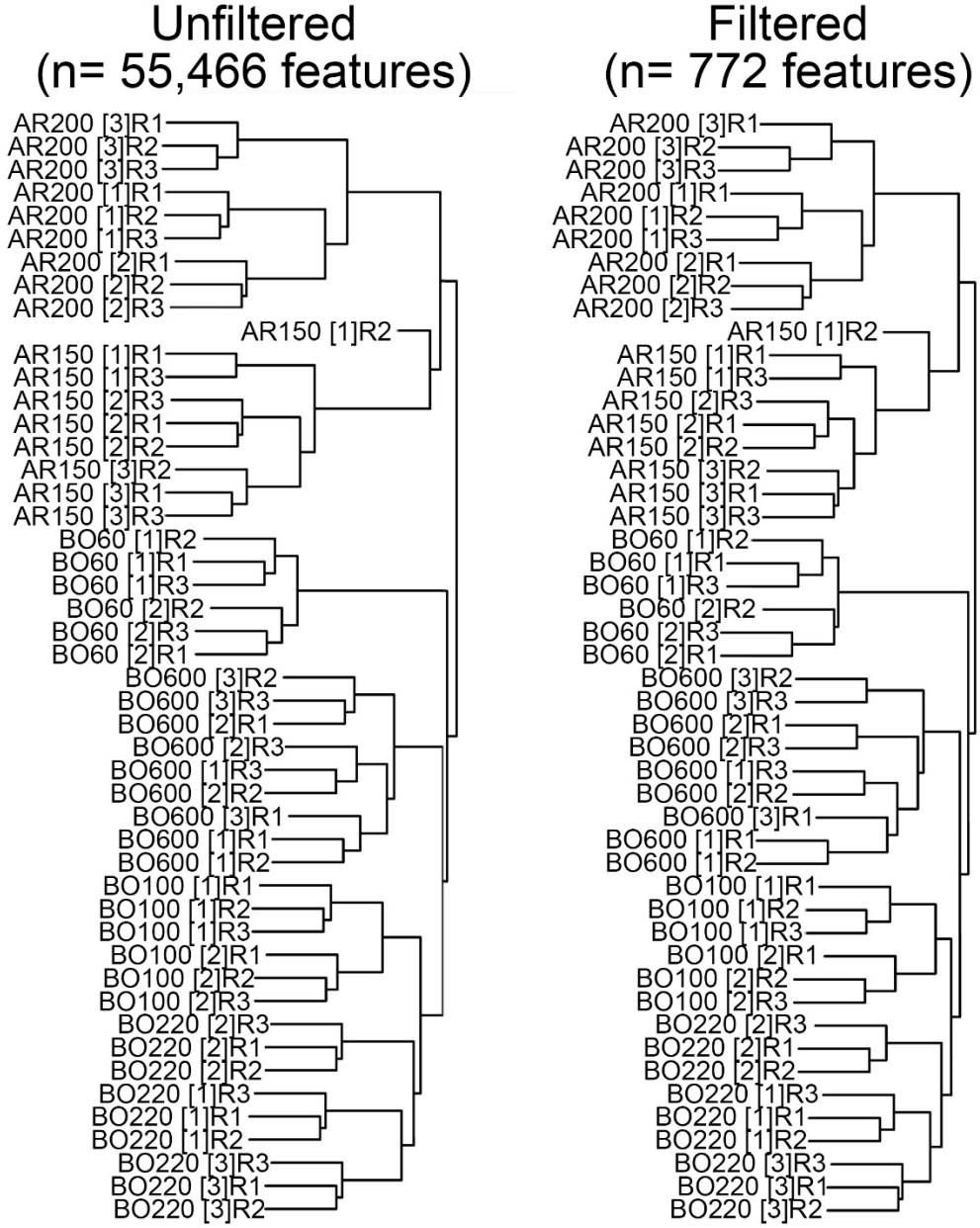


Figure S4.1 IMS-MS spectra for water samples for all time points and conditions.

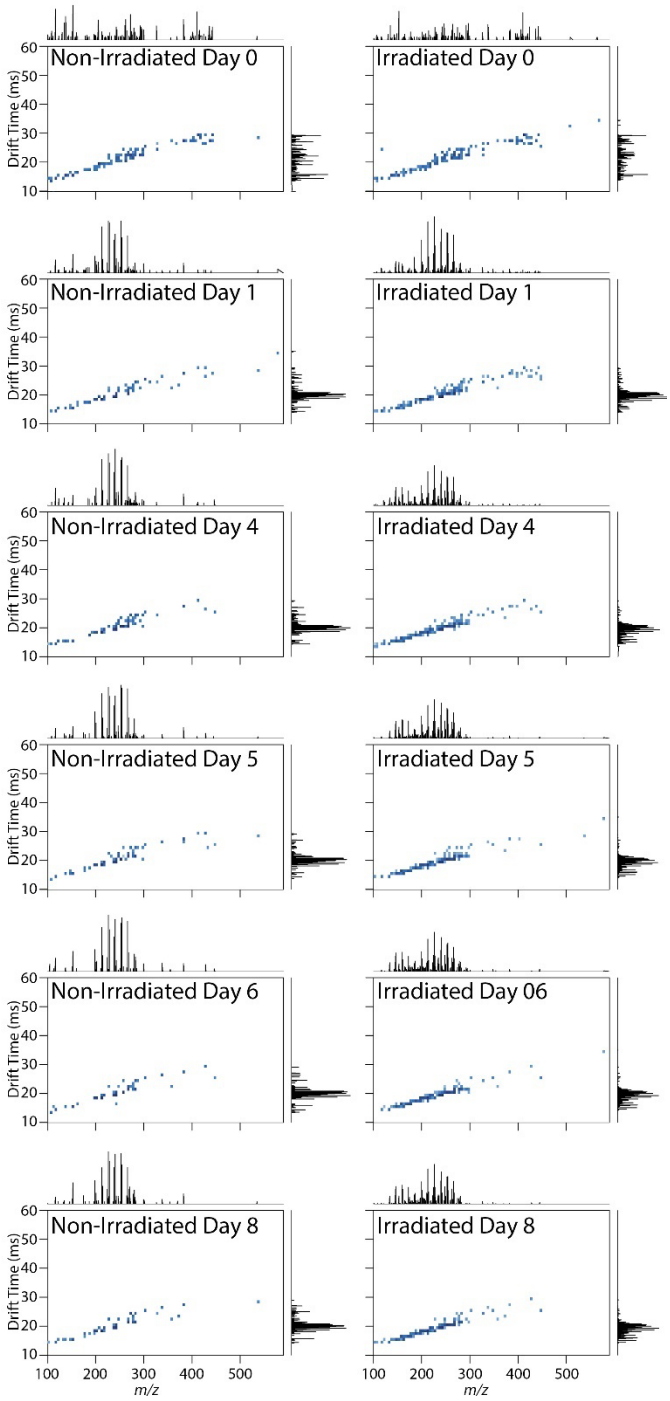
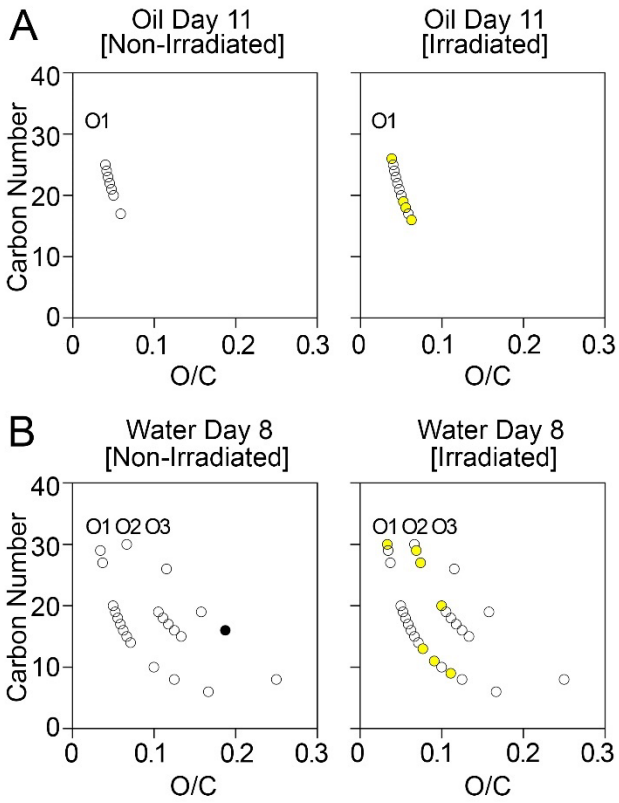


Figure S4.2 O/C versus carbon number plots generated in (A) oil at day 11, or (B) in water at day 8. Molecules are colored according to their presence in each condition (black circles, present in non-irradiated conditions; yellow circles, present in irradiated conditions; white circled, present in both conditions).



APPENDIX B

SUPPLEMENTAL TABLES

Table S2.1. Identifiable information for the samples evaluated in this study. Samples were selected from the Gulf of Mexico onshore and offshore regions. Onshore samples from Texas (TX), Louisiana (LA), and Alabama (AL). Samples from offshore were from High Island (HI), East Cameron (EC), or South Timbalier (ST).

Table S2.2. Compounds characterized through GC-MS, biomarker and IMS-MS analysis with their corresponding abbreviations.

Table S2.3. Peak Areas of Regular Steranes (m/z=217)

Table S2.4. Selected Ratios of Regular Steranes (m/z=217)

Table S2.5. Peak Areas of Regular Triterpanes (m/z=191)

Table S2. 6. Selected Ratios of Triterpanes (m/z=191)

Table S2. 7. Raw Areas of Mono-aromatized Steranes (m/z 253).

Table S2.8. Selected ratios of Mono- and Triaromatized steranes and Methyl phenanthrene indices.

Table S2.9. Gasoline range hydrocarbon data for on-shore crude oil samples analyzed with GC-MS. Sample identifiers are shown above each plot. The bars indicate relative abundance (% abundance) for each feature as compared to the features listed on the x-axis

Table S2.10. Percent composition of alkanes from C3-C32

Table S2.11. GC-MS alkanes data (% fraction of each identified feature).

Table S2.12. Raw areas of the characterized parent and alkyl polycyclic aromatic hydrocarbons.

Table S2.13. Complete data matrix for IMS-MS analysis of the 19 crude oil samples in triplicates.

Table S2. 14. Filtered data matrix for IMS-MS data analysis. Features were filtered based on a frequency of >1, average of triplicates > 1,000 and overall abundance of >1,000. Features that were etected with same mass-to-charge and CCS values, the

abundance of all features was summed. The fully filtered data set, features 1-939 were selected.

Table S2.15. Fowlkes-Mallow (FM) index measurement of hierarchical clusters (Figure 5) to measure the similarity among clusters based on region (n=2) or areas (n=6).

Table S3.1. Basic compositional and date of collection information of the samples.

Table S3.2. GC-MS full scan data for a mass range of 40 to 500 amu and the normalized abundance of the fragmented ions for the sum of 10,127 scans.

Table S3.3A. IMS-MS Run 1 raw data matrix with experimental triplicates.

Table S3.3B. IMS-MS Run 2 raw data matrix with experimental triplicates.

Table S3.3C. IMS-MS Run 3 raw data matrix with experimental triplicates.

Table S3.4A. IMS-MS Run 1 filtered data matrix with experimental triplicates (1,530 features). See Methods for filtering criteria.

Table S3.4B. IMS-MS Run 2 filtered data matrix with experimental triplicates (4,470 features). See Methods for filtering criteria.

Table S3.4C. IMS-MS Run 3 filtered data matrix with experimental triplicates (4,662 features). See Methods for filtering criteria.

Table S3.5. KMD analysis of IMS-MS (run 1 filtered data from Supplemental Table 4A), with molecular formula, carbon number and DBE assignment.

Table S3.6. Normalized averaged abundance for each sample using IMS-MS (run 1, see raw data in Supplemental Table 3A) with their corresponding mean, standard deviation, fold change, p-value and q-value.

Table S3.7A. Correlation values IMS-MS Run 1 of the filtered data (Table S3.4A).

Table S3.7B. Correlation values IMS-MS Run 2 of the filtered data (Table S3.4B).

Table S3.7C. Correlation values IMS-MS Run 3 of the filtered data (Table S3.4A).

Table S3.8. Hydrocarbon block (carbon chain length and molecular class) data matrix for each sample. Shown are relative abundances for each block.

Table S3.9. Heteroatom species composition for all refined petroleum substances tested.

Table S3.10. P_{adj}-values representative of the difference in carbon number, hydrocarbon class and heteroatom composition profiles among the samples of different production cycles.

Table S3.11. The most abundant features (>1% of total feature abundance in each sample) for products AR150, AR200, BO100 and BO220 with their putative molecular formula assignment and name assignment.

Table S4.1. Sample identification and information on exposure and collection.

Table S4.2. Raw IMS-MS data matrix (n = 6,892) of the 42 water samples and solvent blank with their corresponding technical replicates (T.R.).

Table S4.3. IMS-MS data matrix (n = 759) with filtering criteria of abundance greater than 500, instrumental and experimental (Day 0) blanks subtracted and instrumental replicates averaged with molecular characterization based on KMD and discriminating feature clustering information.

Table S4.4. Biomimetic Extraction-Solid Phase Microextraction Gas Chromatography-Flame Ionization Detection measurements [µmol as 2,3 dimethylnaphthalene/mL PDMS] for acidified and poisoned water samples.