HHS Public Access
Author manuscript
Regul Toxicol Pharmacol. Author manuscript; available in PMC 2024 January 01.

Published in final edited form as:
Regul Toxicol Pharmacol. 2023 January ; 137: 105310. doi:10.1016/j.yrtph.2022.105310.

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

Analytical Chemistry Solutions to Hazard Evaluation of Petroleum Refining Products

Alina T. Roman-Hubers[1], **Alexandra C. Cordova**[1], **Mark P. Barrow**[2], **Ivan Rusyn**[1,*]

[1]Interdisciplinary Faculty of Toxicology and Department of Veterinary Physiology and Pharmacology, Texas A&M University, College Station, Texas, USA

[2]Department of Chemistry, University of Warwick, Coventry, CV4 7AL, United Kingdom

Abstract

Products of petroleum refining are substances that are both complex and variable. These substances are produced and distributed in high volumes; therefore, they are heavily scrutinized in terms of their potential hazards and risks. Because of inherent compositional complexity and variability, unique challenges exist in terms of their registration and evaluation. Continued dialogue between the industry and the decision-makers has revolved around the most appropriate approach to fill data gaps and ensure safe use of these substances. One of the challenging topics has been the extent of chemical compositional characterization of products of petroleum refining that may be necessary for substance identification and hazard evaluation. There are several novel analytical methods that can be used for comprehensive characterization of petroleum substances and identification of most abundant constituents. However, translation of the advances in analytical chemistry to regulatory decision-making has not been as evident. Therefore, this review's goal is to bridge the divide between the science of chemical characterization of petroleum and the needs and expectations of the decision-makers. Collectively, mutual appreciation of the regulatory guidance and the realities of what information these new methods can deliver should facilitate the path forward in ensuring safety of the products of petroleum refining.

Introduction

Crude oils are naturally occurring and highly complex substances which vary considerably in molecular composition according to their origins; they comprise a myriad of constituents,

---

[*]Corresponding author: Ivan Rusyn, MD, PhD, Veterinary Physiology and Pharmacology, 4458 TAMU, Texas A&M University, College Station, TX 77843, USA; Tel: +1-979-458-9866; irusyn@tamu.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

CRediT author statement
Ivan Rusyn, Supervision, Conceptualization, Writing - Review & Editing
Alina T. Roman-Hubers, Data Curation, Methodology, Formal analysis, Data Curation Writing - Review & Editing
Alexandra C. Cordova, Methodology, Writing - Review & Editing
Mark P. Barrow, Methodology, Methodology, Writing - Review & Editing

Declaration of Competing Interest
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

primarily hydrocarbons, but also other organic and inorganic molecules (Smith et al., 1959). Close to one hundred billion barrels of crude oils are annually extracted, distributed, and processed into a wide variety of refined petroleum products (Kaiser, 2017; Salvito et al., 2020). The chemical composition of petroleum refining products therefore depends on both the type (*i.e.*, origins) of crude oil from which it was derived, and the refining process (*i.e.*, fractional distillation and/or cracking followed by additional processing through solvent extraction, hydro-desulfurization, or hydrogenation) used to meet performance characteristics of the end-products (McKee et al., 2015). Products of petroleum refining are high production volume substances and thus are heavily scrutinized in terms of their potential human and environmental health hazards and risks. Because of inherent compositional complexity and variability, petroleum substances are prototypical representatives of a diverse class known as substances of unknown, variable composition, complex reaction products, or biological materials (UVCBs); these substances present unique challenges to regulatory agencies, especially in terms of characterization of their chemical composition (Clark et al., 2013; ECHA, 2017c; Lai et al., 2022). It is worth noting that the different fields within the academic, industrial, and regulatory science communities can use differing terminology. Analytical researchers often refer to petroleum refining products as “hydrocarbon mixtures” which are, in turn, part of the broader “complex mixtures” family of samples. The term “mixture” is avoided by the industry and decision-makers as they reason that most substances in commerce that are made from oil are products of refining, rather than mixing, and thus to differentiate from the mixtures found in the environment, such substances are called petroleum UVCBs.

Studies of molecular composition of crude oils and petroleum refining products have a long history spanning over 80 years (Figure 1). The analytical characterization of petroleum substances historically tracked the physico-chemical properties that pertained to the functionality of the product, such as flash points and vapor pressure. With the advent of spectroscopy and mass spectrometry techniques, there came the possibility to gradually gain more detailed understanding of composition; however, the granularity of information on the constituents in registered petroleum products is still lacking. Recent improvements have been made in the resolution of mass spectrometers, ionization methods in order to access a wider range of components, separation methods to offer structural insights, and developments in data visualization through standardized diagrams (Palacio Lozano et al., 2020). As a result of these advances, new opportunities emerged to provide comprehensive characterization of these complex substances and satisfy regulatory needs on the composition, quantity of potentially hazardous constituents, and the extent of variability among manufacturing batches of these products. The range and types of mass spectrometry techniques that can be used for the analysis of petroleum-related samples is quite extensive; collectively, the methods for study of petroleum are now being referred to as “petroleomics,” a sub-field of analytical chemistry aiming to identify the totality of constituents of crude oil and petroleum refining products using high resolution mass spectrometry methods (Hsu et al., 2011; Marshall and Rodgers, 2004; Palacio Lozano et al., 2020).

Despite major advances in the ever-improving analytical resolution of individual molecules and their classes in oils and complex petroleum UVCBs (Wise et al., 2022), there has been relatively little use of the data from these new methods and instruments in regulatory

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

submissions, with the exception of GC×GC-FID-derived data (Redman et al., 2014; Ventura et al., 2011), or even their mention in the reviews or original research publications (Figure 2A). The naming conventions and approaches to identification of complex petroleum UVCBs remain rather imprecise (Rasmussen et al., 1999); only general compositional characteristics are used to define broad manufacturing categories (Salvito et al., 2020). While such information is generally sufficient for naming and identification of petroleum UVCBs (ECHA, 2017a), it is often not sufficient for evaluation of human health and environmental hazards, a prerequisite to registration and authorization for their use (Salvito et al., 2020).

The regulatory science and analytical chemistry fields run in parallel and both are highly specialized, requiring significant expertise. Consequently, intricacies of the legislative mandates governing regulatory decision-making are often unfamiliar to the researchers who develop and refine advanced methods for petroleomics. Similarly, decision-makers may not be aware of the latest opportunities that analytical chemistry has to offer. Collectively, there is a considerable gap in the translation of knowledge from petroleomics-focused analytical laboratories to applied decision-making. This review aims to first summarize the regulatory guidance for characterization of the chemical composition of petroleum UVCBs and then demonstrate how existing petroleomics techniques could be applied to address these needs. Recent additions to the European Union (EU) Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) regulatory guidance for demonstrating the composition of complex UVCB (ECHA, 2022), together with recent advancements in petroleomics applications that are poised to address these needs (Palacio Lozano et al., 2020), create a unique opportunity to bridge the divide. Here, we highlight the opportunities that are already within reach for using modern analytical and data analysis/visualization techniques for impactful decisions on petroleum UVCBs. We reason that it is imperative for decision-makers to be aware of the possibilities and limitations of current analytical approaches so that regulations can be sufficiently strict, yet realistic in terms of their attainment using best available science.

What Information do the Decision-Makers Seek on Petroleum UVCBs?

Most impactful guidance to the industry on petroleum UVCBs has been produced in the United States and the European Union. These include the United States Environmental Protection Agency (US EPA) High Production Volume (HPV) Challenge Program, a voluntary industry-government information sharing effort that was launched in 1998 (Petroleum HPV Testing Group, 2017), and the REACH legislation and associated guidance documents in the European Union (ECHA, 2017c) which is now being adopted in other countries around the globe. Because REACH is the most recent and stringent legislative regime, it has become a *de facto* global driver for the regulatory scrutiny of both new and existing chemicals, including petroleum UVCBs. Therefore, most decision-making contexts discussed herein pertain to EU REACH regulation and its implementation by the European Chemical Agency (ECHA) through guidance documents.

Simply put, REACH-based guidance for UVCBs states that data provided by the registrants shall enable (i) identification of the substances that are submitted for registration, and

Author Manuscript

(ii) evaluation of the potential hazards to human health and the environment (ECHA, 2017c). In both instances, information on chemical composition of a substance is required; however, in a slightly different context (Table 1). For the former, data requirements are typically less stringent because complex substances may be registered based on the manufacturing process, intended use, and/or physical-chemical properties. For the latter, the molecular identification (both elemental composition and structure) data requirements are far greater because the individual constituents that may be present in a "representative" sample, their amounts, and presence of known or suspected hazardous substances must be reasonable ascertained to enable grouping and read-across among petroleum UVCB substances. Based on these two initial components of REACH, an EU regulation that came into force in June of 2007 (European Council, 2007), decisions on authorization of use or restriction(s) are made. An additional challenge for petroleum-derived and other UVCBs is that their composition is inherently variable from batch to batch of the nominally the "*same*" product; this is also true for product-to-product variability within broader categories (FuelsEurope, 2015). This variability may be due to the source crude oil used in manufacturing; petroleomics-related research has revealed significant differences in compositions of petroleum according to its origins, slight variations in manufacturing processes among refineries, and naturally occurring degradation and weathering processes. Because safety testing is typically conducted with a "*representative*" sample of a UVCB product or category, the confidence with which these data can be extrapolated to other samples of the same product or to other products that are "*similar*," especially when sample composition data may be commercially sensitive, depends on how much and what type of data is available to ascertain such similarity.

For the purpose of substance identification, the regulatory frameworks in the United States and the European Union have historically (Figure 1) named, grouped and categorized petroleum UVCBs based on the manufacturing processes used in oil refining, as well as physical-chemical properties and other broad chemical fingerprinting data (Dimitrov et al., 2015; ECHA, 2008; Rasmussen et al., 1999). Manufacturing process-centric naming conventions for these complex substances were originally developed by the American Petroleum Institute (API) and the US EPA for the purpose of creating an inventory of petroleum products under the Toxic Substances Control Act inventory (API, 1983; EPA, 1995; U.S. EPA, 1978). Both Chemical Abstract Service (CAS) and European Inventory of Existing Commercial Chemical Substances (EINECS) identifications have been assigned to a large number of petroleum UVCBs, even though the descriptions of each of these substances under either one of these "unique identifiers" are rather imprecise and far from being unique (Rasmussen et al., 1999). The broad substance categories were somewhat refined under the US EPA's HPV Challenge Program through addition of more detailed information on physical-chemical properties, as well as some human and environmental hazards data (Petroleum HPV Testing Group, 2017).

Far greater information requirements, with respect to both chemical composition and potential hazards, were imposed by REACH (European Council, 2007). Petroleum UVCBs are high production volume substances that were subject to the earliest deadline for registration and the most stringent requirements for hazard evaluation. From 2007 to 2010, about 8,000 registrations were submitted in the EU for petroleum substances that were

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

produced or imported at >100 tons/year (CONCAWE, 2022). Subsequently, the number of registrations of petroleum products was reduced to 191 substances through consolidation of redundant submissions and further grouping of substances deemed to be "similar" based on a variety of considerations. To put the scale of the challenge in context, tens of thousands or even hundreds of thousands of unique molecular formulae can be observed in a single fraction of a petroleum sample using ultrahigh resolution mass spectrometry, indicating that potentially millions of different structures are present in individual crude oils when allowing for isomers (Palacio Lozano et al., 2019b; Palacio Lozano et al., 2020). While the registration submissions for petroleum UVCBs were completed more than 10 years ago and are regularly updated (CONCAWE, 2021), discussions between trade associations and regulatory agencies are ongoing to determine the most sensible ways to improve the dossier quality and ensure the information is compliant with the REACH regulation, primarily by generating more testing information and reinforcing read-across and category approaches. Still, industry's attempts to waive animal testing requirements through read-across have been rejected by ECHA because of considerable data gaps in hazard assessment and compositional characterization (ECHA, 2020a; ECHA, 2020b; ECHA, 2021).

For registration under REACH (Annex VI, Section 2), the data should be sufficient to enable substance identification (Table 1). For petroleum substances, ECHA guidance is that the following data should be provided: (1) accepted nomenclature; (2) appropriate identifiers such as source/feedstock, refining history, boiling and carbon number ranges, physio-chemical characteristics, chromatographic or spectral information, flash point, and viscosity; and (3) compositional information including identification and concentration of the individual constituents present at >10% and that are known to be hazardous, persistent and/or bioaccumulative, identification of any additives, and generic description of unknown constituents (CONCAWE, 2012; ECHA, 2017a). Additionally, Articles 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*" (ECHA, 2017b). Based on these guidance documents, the registrants (companies or trade associations) have traditionally relied on a wide range of analytical techniques (Figure 2A) to furnish the information on petroleum substance identification (Clark et al., 2013; CONCAWE, 2012; CONCAWE, 2014; CONCAWE, 2020).

For the evaluation step, REACH specifies (Annex XI, Section 1.5) that substances may be grouped based on "*structural similarity between substances which results in a likelihood that the substances have similar physicochemical, toxicological and ecotoxicological properties so that the substances may be considered as a group or category.*" Next, prediction of possible hazards of data-poor substances are made through the application of "read-across" from an analogous substance that has been tested in a requisite assay and is deemed "similar." For this step, REACH regulation states that "*it is required that the relevant properties of a substance within the group may be predicted from data for reference substance(s) within the group (read-across approach).*" The registrant shall establish a read-across hypothesis which explains what structural similarities or differences exist between the source and target substance(s) to which read-across is applied and why a prediction for a toxicological or ecotoxicological property can be made with confidence.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Recently, REACH regulation (Annex XI, Section 1.5) has been amended to state that for the application in 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). Once such information becomes available, and in cases where structural differences are present between the source and target substances, the read-across hypothesis should explain why the differences in the chemical structures within a group will not influence the toxicological or ecotoxicological properties or may do so in a regular pattern.

Even though the regulatory language above may seem rather straightforward, in practice the bar on establishing "*structural similarity*" is very elusive in the case of petroleum UVCBs. It is widely acknowledged that the chemical complexity of petroleum substances far exceeds the capabilities of any one method, even the highest resolution mass spectrometers; therefore, a combination of techniques and approaches is often employed. However, despite the considerable scientific advances achieved in the past decade (Figure 1) in both molecular separation (encompassing chromatography, spectrometry and ionization) and detection (various modalities of mass spectrometry, flame ionization detection, and spectroscopy), many challenges with precision and confidence in comprehensive molecular characterization of petroleum samples persist.

For example, a recent study using Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS) revealed nearly a quarter of a million molecular formulae in a fraction from one petroleum sample (Palacio Lozano et al., 2019b). This study is an example of the reality that millions of structures (as opposed to formulae, as structure also determines toxicity) are present in petroleum UVCBs, an intractable challenge both for the chemical analysts to resolve and identify, as well as for the decision-makers to evaluate. Indeed, even the REACH regulation itself (Annex XI, Section 1.5) acknowledges that "*if it can be demonstrated that the identification of all individual constituents is not technically possible or impractical, the structural similarity may be demonstrated by other means, to enable a quantitative and qualitative comparison of the actual composition between substances*" (European Commission, 2021). Clarification regarding the definition of "*other means*" was recently released May 2022 (ECHA, 2022), where specific justification for lack of knowledge about constituents, lack of published methods to identify constituents, and explanation about technical hindrance to resolution and identification of constituents comprising >20% of the substance is necessary to apply such means for characterization. Where "other" means are justified, quantitative comparison of constituents in common between source and target substances should be included, as well as a qualitative comparison of structures that vary between the substances. Fingerprinting, for example, can be used if the following are addressed: information on >95% of all constituents, information on the constituents of high concern, and high analytical resolution to enable accurate alignment, quantitation, variability, and structural data (beyond molecular formulas) of constituents between substances (ECHA, 2022). Therefore, the subsequent sections of this review are framed around three overarching critical needs/questions that REACH regulation challenges the registrants of petroleum UVCBs to address to gain regulatory acceptance of the grouping and read-across hypotheses that are addressable by means of standardized analytical chemistry methods:

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

- Critical Need 1: Providing detailed information on the *structure* of the constituents;
- Critical Need 2: Providing information on the *concentration* of the individual constituents; and
- Critical Need 3: Demonstrating *compositional similarity* of complex petroleum UVCBs through other means when the identification of all individual constituents is not technically possible or impractical.

Conventional Methods for Characterization of Petroleum Substance Identity and Composition

There are many methods for characterization of physical-chemical properties and chemical composition of oil, that have also been applied to petroleum UVCBs (Figure 2A). Multiple techniques are used due to the broad range of substances with widely different composition, volatility, and polarity (CONCAWE, 2014; Stout and Wang, 2007; Wang et al., 2011). Most publications concerning characterization of petroleum UVCBs have explored approaches to define elemental composition, physical properties, and gross structural information using nuclear magnetic resonance (NMR) or infrared spectra of these substances. The elemental analyses for characterization of petroleum UVCBs evaluate the concentrations of major elements ranging from carbon and heteroatoms to metals (CONCAWE, 2019). There are specific methods for assessing physical properties, which typically relate to the quality of the product and are thus more regularly measured during manufacturing and more available. For example, measurements of specific gravity using American Society for Testing and Methods (ASTM) methods such as *ASTM D287 Standard Method for API Gravity of Crude Petroleum and Petroleum Products* are in wide use (Giles, 2016). Other physical-chemical information, such as boiling and carbon number ranges, is typically deduced through physical (*e.g.*, ASTM methods D86, D1160 and D2892), or simulated (CONCAWE, 2019; CONCAWE, 2020) distillation methods.

Spectroscopic techniques have been widely employed to obtain broad compositional information for regulatory characterization and identification of UVCB substances but their utility for the analysis of petroleum UVCBs has been questioned (CONCAWE, 2020). NMR methods (IP392, ASTM D5292) measure the percent of carbon or hydrogen atoms in an aromatic ring (CONCAWE, 2020). Infra-red spectroscopy measures the presence of functional groups to define the degree of saturation in the constituents (CONCAWE, 2012). Ultra-violet spectroscopic analysis quantifies compounds by detecting unsaturated bonds such as those in olefins and aromatics, as well as ketonic and heteroatom groups, but is limited in resolution for other constituents.

More detailed compositional information, which gives greater insight into the chemical classes and carbon ranges of the substance, is obtained using chromatographic techniques that enable separation of constituent groups in complex petroleum UVCBs, these include gas- and liquid-based approaches. Gas chromatography (GC)-based analyses predominate; gas-based analysis of hydrocarbons were first published in the 1960s leading to the development of a standardized method (ASTM 2887–84) for determination of *Boiling*

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Range Distribution of Petroleum Fractions by Gas Chromatography (Giles, 2016). GC is a powerful tool used for the separation and semi-quantitative assessment of non-polar constituents such as hydrocarbons and polycyclic aromatic hydrocarbons (PAH) (CONCAWE, 2012). It offers better separation than liquid chromatography (LC) but is limited by the boiling point of compounds (affecting the accessible mass range), which is a greater hindrance for the characterization of the heavy petroleum that is increasingly relied upon, and sometimes compounds must be derivatized to ensure GC compatibility. LC is dependent on the polarities of the constituents present, predominantly used to characterize less volatile polar compounds. High-performance liquid chromatography (HPLC; e.g., ASTMD6379 and IP391 methods) is used to quantify mono-, di- and tri-aromatic hydrocarbons (CONCAWE, 2012; CONCAWE, 2019). Meanwhile, thin layer or liquid column chromatography (ASTM D2007) separation generates information on basic chemical properties (CONCAWE, 2012; CONCAWE, 2019). Gas chromatography coupled to mass spectrometry detection (GC-MS) is widely used in forensic fingerprinting (US EPA 8270 and 8051B) and to characterize the composition of petroleum UVCBs (US EPA, 1996; US EPA, 2014). Flame ionization detection (FID) has been coupled with both LC and GC for the detection and quantification of hydrocarbons (CONCAWE, 2012).

However, conventional standardized methods detailed above are insufficient to establish truly comprehensive compositional characterization of UVCBs as needed by REACH to accept grouping and read-across hypotheses from registrants. Recent decisions by ECHA on testing proposals provided several reasons as to why that is the case (ECHA, 2020a; ECHA, 2020b; ECHA, 2021). In these decisions, ECHA noted that (i) physical-chemical characterization of whole complex substances does not demonstrate similarity of chemical constituents of these substances; (ii) elemental and other traditional analysis methods do not provide information on the identity and concentration of individual chemical constituents, but rather provide physical-chemical characterization of the substance as a whole; and (iii) standard methods used in the submissions provided insufficient information to estimate the variability of constituents both within and among substances and groups. ECHA deems these criteria necessary to establish the applicability domain of a given category to confirm membership of the source substance(s) and enable subsequent read-across to the target substance(s) (ECHA, 2020a; ECHA, 2020b; ECHA, 2021).

Collectively, despite the use of a battery of analytical assays, expending large sample quantities on some of these analyses, using specialized sample preparation techniques, and incurring considerable costs to acquire these data, the registrants did not establish substance characterization that would be acceptable by ECHA. Indeed, the industry itself acknowledges that spectroscopic techniques are limited to bulk characterization and "*most substances* [*i.e.*, petroleum UVCBs] *cannot be effectively differentiated from each other by UV, IR*, *^{1}H-NMR or ^{13}C-NMR spectroscopies*" (CONCAWE, 2020). Further, even though chromatography-based methods provide considerable amount of information for characterization of nonpolar and relatively volatile compounds (*e.g.*, ASTM D2134, D6729, D6730, among others (CONCAWE, 2012), as well as aliphatic and aromatic fractions (Reddy and Quinn, 1999; Wang and Fingas, 2003), their limited resolution leaves much of the complex substance uncharacterized (Wang et al., 2011; Weng et al., 2015). When assessing the needs for analysis of complex substances, it is important to consider that

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

different analytical approaches are complementary and bring their own advantages and disadvantages. With the limitations of many of the more routine methods, there has been increasing use of ultrahigh resolution mass spectrometry and other advanced methods.

More recently, two-dimensional GC (GC×GC) technique has been applied to petroleum substances because it allows for an even greater separation of the multitude of constituents. Specifically, the coupling of GC×GC with FID allows qualitative constituent information at the level of carbon number and chemical class (ASTM International, 2011); this has been useful for simplifying the composition of petroleum substances by binning molecules by "hydrocarbon block" (Redman et al., 2012). Coupling GC×GC with mass spectrometry provides more structural information on the specific constituents (Jennerwein et al., 2014; Mao et al., 2009).

High and Ultrahigh Resolution Mass Spectrometry Techniques

Advancements in mass spectrometry over the past 80 years (Figure 1) have spawned the application of high-resolution approaches for the study of petroleum substances at a molecular level (Palacio Lozano et al., 2019a; Wise et al., 2022). Resolving power is one of the key performance metrics of any mass spectrometry and is typically defined as $\frac{m}{\Delta m}$, where m is the m/z of the ion of interest and Δm is the width of the peak at half its height, using the full width at half maximum (FWHM) definition. In essence, the higher the resolving power, the more peaks can be observed for complex samples, as this reduces overlap of peaks (Phillips et al., 2022). "High resolution" has typically been accepted to mean a resolving power of >10,000 (Xian et al., 2012), but it has become well-established over the past two decades that "ultrahigh resolution," which often refers to a resolving power of >100,000, is essential for characterization of the most complex samples such as petroleum.

The second performance metric that should always be considered is that of mass accuracy. An instrument that offers high mass accuracy indicates that it typically provides data with low mass errors. When an elemental composition (molecular formula) has been assigned to an observed peak, the observed m/z and the theoretical m/z for the assignment can be used to calculate the mass error, measured in parts per million (ppm), using the following equation:

$$\frac{m_{observed} - m_{theoretical}}{m_{theoretical}} \times 1,000,000$$

Note that a negative mass error indicates the peak appears at a lower m/z than the theoretical value and a positive mass error indicates the peak appears at a higher m/z than the theoretical value; for each peak, the closer the mass error is to zero, the greater the confidence in the given assignment of the molecular formula. It is also important to note that, while researchers desire resolving power to be as high as possible, they strive to keep mass errors as low as possible. Thus, the highest resolving powers afford researchers the ability to observe more components within complex samples, while the highest mass accuracies (lowest mass errors) afford greater confidence in the assignments of elemental composition

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

and structures, together establishing detailed compositional “profiles,” “fingerprints,” or “signatures” for complex samples. Within the field, there has been discussion about the most appropriate terminology to use, drawing parallels with how fingerprints do not normally change but that signatures do, and how this understanding may be related to compositions of complex substances changing when subject to anthropogenic or environmental processes.

The comprehensive characterization of molecular composition of petroleum through high resolution mass spectrometry is an active area of investigation and includes several approaches to precise detection, naming, and structural characterization of the individual constituents (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; Roman-Hubers et al., 2022; Xian et al., 2012). Modern time-of-flight (TOF) mass spectrometers, which are widespread and considered high resolution, offer resolving powers typically in the range of 10,000–60,000. By contrast, FT-ICR MS is the highest performance variety of mass spectrometer and is considered to offer researchers ultrahigh resolution, at one or two orders of magnitude higher performance. FT-ICR MS is based upon ions orbiting inside a cell, which is in turn housed within the bore of a superconducting magnet, and the technique offers ultrahigh resolving power (~10^6 FWHM) and mass accuracy (sub-ppm). Orbitrap mass spectrometers are a newer variety of mass spectrometer based upon a Kingdon trap design, rather than using magnetic fields; these instruments typically offer a resolving power of ~10^5 FWHM. FT-ICR MS and Orbitrap MS offer differing degrees of ultrahigh resolution, with FT-ICR MS offering the highest performance; however, the advantage of TOF MS is in rapid acquisition time which allows for coupling with additional separation techniques such as two-dimensional gas chromatography and ion mobility spectrometry (Palacio Lozano et al., 2019a). These three analytical techniques (Figure 2B) are actively used in petroleomics analyses because they offer somewhat different approaches to determining molecular formulae present in complex substances (Palacio Lozano et al., 2019b; Rodgers and McKenna, 2011), but have not yet been used for detailed petroleum substance characterization for registration or evaluation purposes.

Ultrahigh Resolution MS Data Processing and Visualizations

Petroleum substances contain highly homologous series of hydrocarbon molecules; thus, complex substance analysis can be facilitated by exploiting the patterns of various chemical groups. The method of Kendrick mass defect (KMD) analysis facilitates sorting molecules into homologous series (Kendrick, 1963). The composition of complex petroleum substances can be visualized using the KMD approach because most molecules belong to homologous series comprised of (CH_2) alkyl groups and other functional groups (Figure 3A), this method has been widely used in petroleomics (Hughey et al., 2001; Marshall and Rodgers, 2004; Marshall and Rodgers, 2008; Palacio Lozano et al., 2020). Due to the high resolution and mass accuracy, the molecular composition assigned to the ions that fall in or out of the homologous series can be used to predict their elemental content ($HC_\#$, $O_\#$, $N_\#$, $O_\#$, $S_\#$), carbon number, and double bond equivalents ($DBE = C_\# - H_\#/2 + N_\#/2 + 1$). A number of other visualizations have been proposed to express the molecular composition of various substances based on the rings and double bonds in the carbon framework (*i.e.*, DBE) of the constituents that can be plotted against their carbon number (Figure 3B). Van

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Krevelen diagrams (Figure 3C) are used to display the degree of aromaticity and oxidation of constituents by plotting the H/C versus O/C ratio of the organic compounds in the complex substance (Kim et al., 2003; Van Krevelen, 1950; Van Krevelen, 1984). Relative abundance of various classes of compounds in the samples is typically plotted as a chemical class bar chart (Figure 3D), or as two-dimensional “hydrocarbon blocks” (Figure 3E).

Molecular-level analyses using ultrahigh resolution MS typically generate voluminous datasets even after processing by the software provided by instrument manufacturers. A number of third-party software packages have been developed recently to facilitate data analysis and visualizations. ***Peak-by-Peak*** fee-based software (Spectroswiss, 2019) is tailored for processing FTMS raw data on transients and mass spectra and generates output files for follow-up statistical analyses. ***PetroOrg*** (Riches et al., 2015) and ***Composer*** (Sierra Analytics, 2022), launched in 2014 and 2008, respectively, are two stand-alone fee-based packages that allow processing and visualization of the molecular composition of petroleum substances data acquired through high resolution MS. ***UltraMassExplorer*** is an open source web-based package that uses *R Studio* (Leefmann et al., 2019). It applies static formula libraries for molecular formula assignment of the molecular formulas based on neutral masses coupled to PubChem searches for putative structural assignment. Data visualization is enabled by van Krevelen, KMD and DBE plots. ***DropMS*** is another web-based tool that facilitates high resolution MS data processing and molecular assignments with its corresponding DBE, error, signal intensity, as well as a number of visualizations (Rosa et al., 2020). ***KairosMS*** utilizes an R Shiny interface to process complex data sets produced through hyphenated MS experiments that, when combined with software for formula assignments, can be used for visualization, comparison, and statistical analyses for both direct infusion and hyphenated data set, using a wide variety of approaches (Gavard et al., 2020). ***Interactive van Krevelen*** (Kew et al., 2017) and ***Open van Krevelen*** (Brockman et al., 2018) are two packages that offer interactive diagrams for molecular-level exploration of the data from high resolution MS.

The Regulatory Needs for Characterizing Chemical Composition of Petroleum UVCBs

General considerations and examples of the application of high resolution MS for the analysis of petroleum substances have been reviewed elsewhere (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). However, these previous reviews did not specifically place the application of these techniques in the context of the regulatory needs for registration and evaluation of petroleum UVCBs. Comprehensive characterization, or at least more detailed information on some specific types of constituents, of complex substances based on high resolving power and mass accuracy of FT-ICR MS, Orbitrap MS and TOF MS affords the opportunity to attain valuable information on both molecular composition and relative abundance of the constituents. Such data can also be used to quantitatively evaluate both variability within a substance or group of substances, and similarity between substances. Then, it should be possible to conduct read-across and complete hazard characterization. To provide specific examples of how each technique can

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

be used to address specific regulatory needs, we performed a systematic literature search (with end date of December 2021) focused on the application of each technique to petroleum substances (Figure 4). We found that even though there are many dozens to hundreds of publications on each technique in general, and on their use to analyze petroleum samples in particular, few studies presented data or drew conclusions in a manner relevant to address each regulatory need. The Web of Science search terms and results of the literature search are included in Supplemental Table 1.

Critical Need 1: Providing detailed information on the structure of the constituents.

High mass measurement accuracy and resolving power of modern mass spectrometry analyzers are key steps to address this regulatory need for petroleum UVCBs. Regulatory agencies have repeatedly stressed that "*broad*" compositional information is insufficient to justify groupings and support read-across hypotheses (ECHA, 2014; ECHA, 2020a). While many publications tout ever increasing resolution as an advance in the science of analytical chemistry, we highlight recent examples of studies that focused on comprehensive structural information of the individual constituents in petroleum samples.

With the advances in the performance of mass spectrometers and development of new methodologies, significant strides have been made with respect to characterization of mixtures and petroleum samples. For the analysis of whole crude oils and refined products, FT-ICR MS could resolve hundreds of molecular formulae in the late 1990s (Rodgers et al., 1998) and several thousand in the early 2000s (Hughey et al., 2002; Qian et al., 2001). By 2019, newer FT-ICR MS instrumentation and experiment design led to the identification of ~245,000 molecular formulae in a non-distillable fraction of the maltenes from a heavy petroleum sample, using a resolving power of more than 3,000,000 across the entire m/z range (Palacio Lozano et al., 2019b). Thus, in less than two decades, there was a 60-fold increase in resolving power and 80-fold increase in the number of molecular formulae being assigned within a single sample.

While ultrahigh resolution mass spectrometry has offered unprecedented levels of insight into highly complex samples, such instrumentation typically affords molecular formulae but does not lead to definitive identification of structures. Experimental parameters, such as polarity and ionization behavior, can yield information about probable functionalities of the components, but true structural identification would require additional data. This could be obtained by fragmentation patterns associated with individual peaks (and hence molecular formulae), acquired during "tandem mass spectrometry" (also known as "MS/MS") experiments. As complex samples may comprise tens of thousands or even hundreds of thousands of molecular formulae, performing hundreds of thousands of tandem mass spectrometry experiments to target all of the peaks is not viable, due to the time required and workload due to data analysis.

Combination with an orthogonal separation method, such as a form of chromatography or ion mobility, is an alternative approach for accessing information about functional groups and isomeric contributions. Optimization of these experiments (such as choice of columns, temperature program, etc.) is similarly non-trivial, however, and resolution of components in complex substances can be challenging; the concept of "unresolved complex

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

mixtures”, also referred to as an “UCM hump” on the chromatograms, is well-known (Gough and Rowland, 1990). Furthermore, while chromatography is a well-known method for distinguishing isomers, this is typically combined with a need for authentic standards (e.g., to determine expected retention times on the chromatography column) and databases, although certainly not all available compounds are found within databases. Where there may be dozens of isomers per molecular formula, this would in practice mean a requirement for millions of authentic standards to be run in order to address the most complex petroleum samples that have been characterized by ultrahigh mass spectrometry. This is, again, impractical due to the amount of time required. It is also worth drawing attention to a subtle distinction between molecular formulae (e.g., $C_{18}H_{36}O_2$) and the concept of “peaks;” a peak may represent a single molecular formula in a direct infusion experiment by mass spectrometry (and data is represented by m/z and intensity only), but in combination with mass spectrometry with chromatography, there will be an additional dimension to the data (m/z and intensity, but now also time) and multiple peaks per molecular formula may occur due to the presence of isomers. Combining mass spectrometry with chromatography can yield more peaks by counting of isomers, but not necessarily more molecular formulae.

The isobaric information afforded by ultrahigh resolution direct injection FT-ICR MS has played a significant role in comprehensive understanding of petroleum substance composition; however, the presence of a large number of isomeric hydrocarbons in petroleum substances challenges the utility of ultrahigh resolution methods that do not provide other means for separation of isomeric species. Some studies addressed a growing need to define the isomeric composition of petroleum substances by coupling FT-ICR MS with separation techniques. Gas chromatography prior to FT-ICR MS was used for separation of isomers in petroleum by resolving individual constituents (Barrow et al., 2014; Palacio Lozano et al., 2022). Information of the structural composition and the isomeric diversity can be also attained by coupling with trapped wave ion mobility spectrometry (TWIMS), a variety of ion mobility spectrometry, to FT-ICR MS (Maillard et al., 2021). Multidimensional separation with ultrahigh resolution allows for characterization of individual constituents with accurate mass measurements and structural features. Additionally, high mass accuracy and ultrahigh resolution measurements have elucidated structural information of the individual constituents through post-instrumental data analysis (Cho et al., 2011; Hu et al., 2018).

A number of studies have coupled chromatography to Orbitrap MS for analysis of complex environmental samples (MacLennan et al., 2018; Pereira et al., 2013; Sorensen et al., 2019; Yang et al., 2019). Some studies applied this technique for the analysis of oils or fractions thereof. A combination of GC separation with different ionization methods for Orbitrap MS was used to study volatile components of a petroleum refining product, gas condensate. The authors showed that separation of different isomeric compounds could be achieved using this hyphenated method thus aiding in deeper characterization of a complex substance (Kondyli and Schrader, 2019). More recently, a reverse-phase liquid chromatography method was applied to the analysis of petroleum refining-derived UVCBs and compared to direct injection Orbitrap and FT-ICR MS (Xia et al., 2021). The authors showed that not only could they obtain elemental formulas for a large number of hydrocarbon and heteroatom species, but chromatography-informed retention patterns could also be used to distinguish

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

among isomeric species, hence increasing confidence in structural identification of the individual constituents in petroleum substances.

TOF MS detection, despite its lower mass resolution as compared to FT-ICR and Orbitrap MS, affords an advantage of rapid data acquisition; this technique has been explored extensively by coupling with chromatography or ion mobility spectrometry separations (Palacio Lozano et al., 2019a). The most common type of front-end chromatography in petroleum substance analysis is GC×GC. The additional separation afforded by this technique offers substantial improvement for resolving isomeric constituents and compounds that would otherwise coelute (Ball and Aluwihare, 2014; Luna et al., 2014a; Ngo et al., 2012; Rowland et al., 2011). For example, this technique was used to identify a large number of isomeric molecules, both aliphatic and aromatic, that could not be resolved by a typical GC-MS technique in a complex milieu of hydrocarbons containing a wide (C_{12}-C_{36}) range of carbon numbers (Alam et al., 2016). Similarly, the ability GC×GC-TOF MS to resolve isomers of classical and sulfur-containing naphthenic acids in oil-contaminated environmental samples enabled contamination-source fingerprinting (Bowman et al., 2019) and identification of potentially hazardous substances (Bowman et al., 2020). TOF MS has also been coupled with ultrahigh pressure liquid chromatography to isolate isomeric species in petroleum derivatives (Lv et al., 2013; Mahmoud and Dabek-Zlotorzynska, 2018).

The compositional characterization of petroleum samples has typically relied upon either the use of chromatography or, more recently, the use of ultrahigh resolution mass spectrometry. The separation on the basis of retention time using gas chromatography can be sufficient for monitoring targeted components (*e.g.*, steranes for differentiation of sample origins) and revealing the most significant contributions. As samples become more complex, however, particularly with the increasing use of heavy petroleum sources, there is the increasing probability of coeluting components, as mentioned above, where compounds have the same retention time and therefore cannot be distinguished by one-dimensional chromatography alone. The use of a second column of a different type to result in a two-dimensional approach, GC×GC, can significantly improve the separation and therefore the number of compounds observed.

Due to the temperature ranges of the ovens, gas chromatography-based methods have limited retention times that are accessible, as these are, in turn, linked to the boiling points of the individual compounds. Direct infusion methods coupled with ultrahigh resolution mass spectrometry (*i.e.*, where samples are injected directly into an ion source, with no preceding chromatography) do not have the limitations associated with boiling point, and ultrahigh resolution approaches with direct injection delivery have become increasingly significant, especially for heavy petroleum samples.

There has also been limited coupling of chromatography with ultrahigh resolution mass spectrometry, which affords researchers the ability to separate compounds on the basis of two different dimensions: retention time and m/z. In this way, co-eluting compounds (*i.e.*, the same retention time associated with the GC column) can be separated due to the additional use of the m/z dimension (Barrow et al., 2014). The data sets acquired can be large (*e.g.*, ~25–50 GB) and therefore present data processing and data analysis

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

challenges. The combination of orthogonal approaches, yielding so-called “hyphenated” techniques, represents means by which to access the advantages of the individual methods. Ultrahigh resolution mass spectrometry affords the ability to resolve a greater number of peaks and assign many thousands of unique molecular formulae with confidence, while chromatography or ion mobility affords the ability to separate isomers (same molecular formulae and therefore same m/z, but differing structures due to different arrangements of the atoms). While GC and low resolution GC-MS methods have seen widespread usage for many years, it can be expected that multidimensional GC and the coupling of orthogonal methods with ultrahigh resolution mass spectrometry will both become increasingly used.

Obtaining structural isomeric information is also possible using ion mobility mass spectrometry (IMMS), a post-ionization separation technique, often coupled to TOF MS and used for the analysis of petroleum samples (Santos et al., 2015). IMMS provides structural information that is complementary to the observed m/z of a compound by characterizing the spatial conformation of individual constituents via their drift time through an inert gas and the subsequent derivation of collision cross section (CCS) values (Dodds and Baker, 2019). CCS values can also be compared with those determined through computational means to determine structures for observed compounds. Currently, this is laborious for complex samples and so not viable for samples with, for example, tens of thousands of molecular formulae and their associated isomers. The combination of ion mobility, mass spectrometry, and computational chemistry does, however, hold potential for providing greater structural insights during characterization of samples. IMMS allows for separation of isomeric compounds by their structural composition (Hoskins et al., 2011; Lalli et al., 2015; Lalli et al., 2017; Mahmoud and Dabek-Zlotorzynska, 2018). The high mass accuracy and resolution of TOF MS, coupled with structural characterization using ion mobility, has allowed comprehensive elucidation of the composition of complex petroleum substances and byproducts (Lalli et al., 2015; Lalli et al., 2017). Confident molecular formula assignment to the IMMS-derived features in a gasoline standard and a crude oil sample was demonstrated through the use of KMD analyses based on CH_2 and H functional units (Roman-Hubers et al., 2021).

While instrumentation has increased in performance, the data analysis methods have often struggled to keep pace. It is evident that highly complex samples are being characterized better than ever before. The field is acutely aware that handling the increasingly complex data is already challenging, but that also there is a need to obtain greater structural insights, going beyond molecular formulae alone and/or ensuring all of the many thousands of peaks are associated with a definitive structure, not only those that are mostly easily targeted. While the progress with complex substance analysis has been remarkable in recent years, there remain mountains to climb.

Critical Need 2: Providing information on the concentration of the individual constituents.

Quantification of individual constituents in complex and multi-constituent substances is a required step for hazard evaluation to ensure no underestimation of the potential human health and environmental hazards (CONCAWE, 2012; ECHA, 2017a). Thus, the ability of high resolution mass spectrometers to determine the thousands of individual constituents is

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

not sufficient without determination of their abundance. Traditional quantitative approaches relying on mass spectrometry detection require the use of various extraction and detection standards, preferably isotopically-labelled ones (Urban, 2016). However, because of the complexity of petroleum substances, the use of standards for absolute quantitation of numerous, rather than a small number of targeted constituents, would potentially mean the need for millions of authentic standards, which is both impractical and impossible. Instead of quantifying absolute concentrations of multiple individual constituents, traditional analysis techniques such as GC-MS and GC×GC-FID derive relative (*e.g.*, fraction of total) amounts for groups of compounds based on a limited number of standards for key classes of constituents. For example, quantitation of individual n-alkanes, selected isoprenoids, polycyclic aromatic and alkyl polycyclic aromatic hydrocarbons, and biomarker compounds is possible by GC-MS (Wang et al., 1994); however, the “UCM hump” of high molecular weight hydrocarbons limits the utility of this technique for hazard evaluation. GC×GC-FID technique is also commonly used for petroleum analyses to derive “hydrocarbon blocks” (ASTM International, 2011). This technique is more amenable for hazard evaluation as it can separate polycyclic compounds with known or suspected hazardous properties (Bierkens and Geerts, 2014); however, this technique is not considered sufficiently informative by some decision-makers, especially for the substances that contain C30 or greater hydrocarbons (ECHA, 2020a). Therefore, there is a great need to determine the ability of high resolution MS techniques to provide quantitation of the individual constituents in petroleum substances.

There are significant challenges with respect to quantifying the abundance of constituents in petroleum UVCBs. The signal observed for a given compound will be influenced by a number of experimental parameters. One of these is, of course, concentration. Other factors include solubility in the chosen solvents, pH of the sample solution, the suitability of ionization method chosen, the polarity of the ion source (*e.g.*, when using electrospray ionization, acidic species will be observed in negative-ion mode while basic species will be observed in positive-ion mode), and instrument tuning, amongst other variables. As one example, alkanes can represent as much as half or more of the composition of some petroleum samples and yet if performing analysis using electrospray ionization, a widespread ionization method which is suitable for observation of polar and ionic species, then the alkanes would essentially go unobserved.

The matter of quantification is also closely linked to structure, as previously discussed. Different isomers of a molecular formula may have differing functional groups which, in turn, may influence the ionization response of the compounds. For instance, an organic molecule containing two oxygen atoms may incorporate two hydroxyl groups or a carboxylic acid, where the carboxylic acid would ionize much more readily than the former, and therefore give a much stronger signal. For this reason, quantification typically involves the coupling of chromatography to mass spectrometry, in order that isomers are separated the signals associated with each isomer can be measured and considered separately. Such experiments are labor intensive, however. Authentic standards must be used for each isomer of each molecular formula, as mentioned previously, but the standards must also be prepared as a series of sample solutions spanning a range of concentrations to measure the signal intensity as a function of concentration, leading to calibration curves.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

The needs for authentic standards and for calculation of the calibration curves leads to increases in researcher workload akin to orders of magnitude. “Untargeted” analyses could necessitate the need for potentially millions of experiments for quantification purposes, depending on the objectives. It is much more common that “targeted” analyses are instead used, where a short list of compounds of concern are searched for, such as benzene, toluene, ethylbenzene, and xylene compounds. Many advanced mass spectrometry approaches which are commonly used are accepted to be semi-quantitative, balancing sample complexity, experimental design, data processing, and time.

Most studies that use FT-ICR MS for the analysis of petroleum samples report tens to hundreds of thousands of detectable constituents; however, this technique is semi-quantitative where abundance of each molecule would depend on multiple factors (*e.g.*, solubility, concentration, ionization response, tuning, etc.). FT-ICR MS studies typically report relative abundances of various hydrocarbon or heteroatom classes rather than that of individual constituents (Bae et al., 2010; Chen et al., 2012; Jennerwein et al., 2014; Kim et al., 2015; Oldenburg et al., 2014; Walters et al., 2015). Some publications focused on detection and quantitation of specific constituents, such as organic sulfur compounds (Lu et al., 2013), or metalloporphyrin complexes (Cho et al., 2014); however, these constituents have uncertain relevance for the purpose of hazard evaluation. The ultrahigh resolution Orbitrap MS technique has not been used for the characterization of the individual constituents; instead, the abundances of the detected ions with inferred elemental composition are used for semi-quantitative evaluation of various fractions and broad chemical classes (Castiblanco et al., 2020; Liu et al., 2020; Rodrigues Covas et al., 2020; Vanini et al., 2020).

Similarly, quantification of the individual constituents using TOF MS is challenging and most studies focused on quantifying abundances of broad classes of compounds (Scarlett et al., 2008). A combination of GC×GC-TOF MS and GC×GC-FID was used to conduct qualitative and quantitative analysis of polycyclic aromatic hydrocarbons for several petroleum UVCBs; however, quantitation was performed only for the constituents for which chemical standards were used (Ristic et al., 2018). Even though absolute quantitation of the individual constituents in complex petroleum samples may be unattainable, the combination of confident molecular formula assignments aided by structural information provided by IMMS and data on relative abundance does enable quantitative evaluation of the most abundant constituents (Roman-Hubers et al., 2021). For example, a study of variability in chemical composition of petroleum UVCBs used the average relative abundance of each constituent in a product across production cycles to determine what molecules may be present at relatively high (*e.g.*, REACH threshold of concern at 0.1% (ECHA, 2017b)) amounts and whether those molecules vary significantly between production cycles (Roman-Hubers et al., 2022). This study is an example of how high resolution MS can be used to not only characterize individual constituents, but also to determine their abundance for consideration as potential substances of concern in hazard evaluation. Indeed, the ability to quantify, even in relative terms, the abundance of the identifiable constituents that comprise most of the petroleum sample is of utmost interest under the REACH framework (Table 1). The most recent advice from ECHA states that “*all constituents present in a concentration at or above 1% must be identified*” when grouping or read-across is proposed

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

(ECHA, 2022). The bar is even greater for specific constituents that may possess hazardous properties, *i.e.*, "*0.1% for constituents that are classified as carcinogenic or mutagenic and 0.3% for substances that are toxic to reproduction or development*." Collectively, "*identified constituents above the thresholds given above must account for a minimum of 80% of the mass of a UVCB substance*" (ECHA, 2022).

Critical Need 3: Demonstrating compositional similarity of complex petroleum UVCBs through other means when the identification of all individual constituents is not technically possible or impractical.

Even though challenges remain in the confident identification of structures for individual constituents in complex substances, the multi-dimensional data from high resolution MS analyses (FT-ICR, Orbitrap and TOF MS) is highly valuable for evaluating broad similarities among complex petroleum UVCB substances and for identifying the degree of variability within a class of substances or between production batches of the same substance (Figure 4). Indeed, a number of studies have used these data to perform statistical analyses and visualize the relationships between samples in large datasets to demonstrate that even broad molecular compositional data can be effective as a means for screening, evaluating similarity and determining what constituents may be variable to prioritize selected samples and constituents that may warrant further targeted quantitative analyses.

The variation in composition of hydrocarbons and heteroatoms resolved using FT-ICR MS has been explored for the analysis of crude oils (Hosseini et al., 2021; Rocha et al., 2018; Silva et al., 2020) and refining products (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), including studies that used such data for categorization of new products (Abib et al., 2020; Hourani et al., 2013; Liu et al., 2014). Through detailed chemical profiles, compositional characterization of different environmental samples allowed detection of the asphaltenes (Neumann et al., 2021; Ruger et al., 2015), fulvic acids (Stenson et al., 2003), oil sands process-affected water (Barrow et al., 2016), bitumen (Lacroix-Andrivet et al., 2021), biochar-derived organic matter (Li et al., 2022a), as well as soil and sedimentary organic matter (Zhong et al., 2011). Studies of oil weathering (Wozniak et al., 2019) and transformation in the environment (Jaggi et al., 2019; Li et al., 2022b; Wozniak et al., 2019) enabled to not only determine the trends in compositional changes, but also to group samples, allow for forensic identification of the related samples, predict the potential environmental impact of oil spills, and designing mitigation strategies.

While there will be limitations to the lowest m/z that can be detected by FT-ICR MS, based upon the magnetic field strength and highest frequency that can be detected (*e.g.*, ~m/z 37 for a modern 12 T instrument), with appropriate instrument tuning, both Orbitrap and FT-ICR mass spectrometers can offer comprehensive insights into the lower molecular weight compounds of the chemical profile of complex substances, not afforded by other high resolution MS instruments (Chen et al., 2018; Cheng and Hous, 2021; Cho et al., 2017; Headley et al., 2011). Through the application of Orbitrap MS the variations of the mass spectrum profiles can be traced at a molecular level with low mass error to determine the

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

overall chemical composition (Castiblanco et al., 2020; Dong et al., 2019; Porto et al., 2019; Silva et al., 2019; Vanini et al., 2020). Liquid chromatography separation techniques coupled to Orbitrap MS have provided qualitative and quantitative monitoring of the organic species through the chemical profile complex substances. (Folkerts et al., 2019; Miles et al., 2020; Sorensen et al., 2019; Xia et al., 2021)

When designing future strategies for screening petroleum-related compounds, it is important to consider what is being measured by different approaches. For example, with electrospray ionization, acidic species would typically deprotonate and be observed using negative-ion mode, while basic species would protonate and be observed using positive-ion mode; acidic and basic species would not be observed in the same experiment, therefore. While this might initially be considered a disadvantage because of the need for two experiments instead of one, on the other hand this can be used to the researcher’s advantage when needing to differentiate by functionality. An organic molecule with a heteroatom content of only one nitrogen atom (*e.g.*, $C_cH_hN_1$) could be pyrrolic (weakly acidic) if observed in negative-ion mode, but would more likely be pyridinic (basic) if observed in positive-ion mode. Where electrospray ionization is suitable for polar and ionic compounds, it does not readily ionize non-polar species and so, for these, an alternative ionization method, such as atmospheric pressure photoionization, would be used. To cover acidic, basic, and non-polar species in an untargeted manner, three experiments may be used, but the number of methods employed could be reduced if instead adopting a targeted approach. Through an awareness of the advantages and disadvantages of the different ionization methods, combined with clear objectives of the screening process (which could become of a targeted nature), it is possible to tailor a system which balances required information and workload.

Multidimensional GC separation coupled with high resolution detection in TOF MS has been widely employed to map out the composition of complex substances. The high-throughput acquisition has been readily employed to screen for saturate and aromatic hydrocarbons to characterize the broad compositional profile (Haitao et al., 2013; Hao et al., 2017; Kulkarni and Thies, 2012; Luo et al., 2016; Qian et al., 2004; Rui et al., 2012). Nevertheless, to achieve a better compositional information and distribution patterns of complex volatile constituents, TOF MS has been coupled with gas chromatography for qualitative and quantitative assessment (Haitao et al., 2013; Hao et al., 2017). When coupled with an HPLC, non-volatile components can be readily separated based on isomeric and isobaric information for comprehensive high resolution characterization (Cao et al., 2020). The drawbacks observed from the above addressed separation techniques prompted the coupling of capillary electrophoresis with TOF MS for qualitative assessment of high molecular weight constituents in a heavy gas oil (Nolte et al., 2013). Compared to other TOF MS hyphenations, GC×GC offers multidimensional high resolution characterization to map out the composition of complex substances (Alam et al., 2018; Frenzel et al., 2010; Gabetti et al., 2021; Luna et al., 2014b; Muller et al., 2020; O’Reilly et al., 2019; Qian and Wang, 2019; Ristic et al., 2018; Zhu et al., 2020). The plots generated from GC×GC-TOF MS analysis plotting constituents in a two-dimensional space (1^{st} retention time versus 2^{nd} retention time) help elucidate the composition of complex substances and facilitates the molecular classification of compounds (Damasceno et al., 2014). Notably, physical characteristics (*i.e.*, density) can be directly correlated based on the detailed chemical composition defined through GC×GC-TOF MS (Vozka et al., 2019).

Petroleum UVCB composition: A regulatory challenge, but what are the solutions?

As summarized above, novel high resolution mass spectrometers offer comprehensive characterization of petroleum substances to qualitatively assess and measure the broad chemical composition of a substance and its individual constituents. The information available through high resolution characterization of the chemical composition of UVCBs 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 constituents which may be used to infer human health hazard properties of the whole substance.

Still, complex UVCBs, especially those produced by refining of oil, remain to be an evolving competency in regulatory science. In the recent past, both decision-makers and industry relied on rather imprecise substance categorization and read-across hypotheses to predict toxicological properties (Clark et al., 2013; McKee et al., 2015; Salvito et al., 2020). Recognizing the lack of clarity in the original regulations and guidelines (ECHA, 2017c), and persistent issues identified in regulatory submissions of petroleum UVCBs (ECHA, 2020a; ECHA, 2020b; ECHA, 2021), additional advice was recently provided with respect to the information on chemical composition for UVCB substances (ECHA, 2022). Indeed, there appears to be a major gap in what information is perceived as sufficient in terms of chemical characterization of petroleum UVCBs by either decision-makers or industry. At the same time, the science of analytical chemistry delivered a number of major advances in terms of novel techniques and visualizations to aid in deep characterization of complex petroleum UVCBs. Scores of research studies have been published, reviews written, and lectures delivered. A "meeting in the middle" is required between those working in regulatory and analytical disciplines, so that the translation of petroleomics science into the practice of decision making on petroleum UVCBs can begin to be realized.

One barrier to such a translation could be the difference of opinions on what "science" is "evidence" and whether the scientific data are "sufficient" to satisfy the regulatory requirements for making decisions (National Research Council, 2009). What is sufficient for a peer-reviewed scholarly publication in a specialized scientific journal may not be sufficient for a regulatory decision. For example, in the EU, it is generally noted that "*available scientific and technical data*" shall be taken into account "*in preparing its policy on the environment*" (Allio et al., 2006). It was noted that this legislative mandate does not call for the available data to be of "best available" quality, even though some of the agencies are required to "*provide the EU institutions and the Member States with the best possible scientific opinions*" (Allio et al., 2006). In the US, the concept of "best available science" has been more defined as a statutory requirement for risk evaluation of chemical substances

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(US EPA, 2017). Specifically, the statute defines *best available science* as "*science that is reliable and unbiased. Use of best available science involves the use of supporting studies conducted in accordance with sound and objective science practices, including, when available, peer reviewed science and supporting studies and data collected by accepted methods or best available methods (if the reliability of the method and the nature of the decision justifies use of the data). Additionally, EPA will consider as applicable: (1) The extent to which the scientific information, technical procedures, measures, methods, protocols, methodologies, or models employed to generate the information are reasonable for and consistent with the intended use of the information; (2) The extent to which the information is relevant for the Administrator's use in making a decision about a chemical substance or mixture; (3) The degree of clarity and completeness with which the data, assumptions, methods, quality assurance, and analyses employed to generate the information are documented; (4) The extent to which the variability and uncertainty in the information, or in the procedures, measures, methods, protocols, methodologies, or models, are evaluated and characterized; and (5) The extent of independent verification or peer review of the information or of the procedures, measures, methods, protocols, methodologies or models.*" In principle, all of these can apply to characterization of the chemical composition of petroleum UVCBs when data are submitted to the authorities for a decision.

Related to the concept of *best available science* is the concept of "*reasonably available information"* which means information that the agency "*possesses or can reasonably generate, obtain, and synthesize for use in risk evaluations, considering the deadlines […] for completing such evaluation*" (US EPA, 2017). To relate the latter concept to the data on chemical composition of petroleum UVCBs, one may argue that a range of technically valid analytical methods applicable to the challenge of characterizing their composition and variability are already available and the data could be generated in a reasonable period of time. There is a balance between cost, availability, time, and the level of information required; the most advanced techniques which ultimately yield the fullest possible understanding may not be suitable for routine screening, whilst more common techniques may afford incomplete information or increased risk of misinterpretation. In this regard, novel analytical advances, such as those described herein, have greatly increased our ability to access valuable information on the composition of UVCBs at a molecular level. At the same time, the advances have revealed in much greater detail the compositional complexity of petroleum and areas where development is still required in order to fully access the range of components and, most significantly, establish structures. Still, strong arguments have been made about the impractical and potentially unnecessary regulatory requirements to deconvolute every possible constituent in a UVCB, even if at some arbitrary abundance cut-off level(s) (Table 1), and that hazard assessment may be sufficiently informed by using bulk compositional data and hydrocarbon blocks (Redman et al., 2012; Salvito et al., 2020). Clearly, the fields of analytical chemistry and regulatory science have not always worked in tandem.

To bridge this chasm, all sides will need to meet in the middle to establish baseline requirements with an understanding of what is analytically possible (Figure 4), including considerations of instrumentation availability, cost and time required for the analysis and data processing. It has been well-documented that different analytical methods, when

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

applied to petroleum substances, have different windows of applicability with respect to the types of molecules they may ionize and/or detect, and may suffer from being semi-quantitative (Aeppli, 2022; Fernandez-Lima et al., 2009; McKenna et al., 2013; Prince and Walters, 2022; Rodgers and McKenna, 2011). Indeed, Figure 4 illustrates the point that novel high resolution analytical methods (FT-ICR, Orbitrap-MS and TOF-MS) have been already applied by a number of authors to the research questions that are directly relevant to the specific regulatory needs in petroleum UVCB space. However, the number of such publications remains very small as compared to studies of the traditional analytical approaches (Figure 2), and seldom do the authors acknowledge the potential for their work to be applied in the regulatory context(s). It is clear that the analytical chemists could better appreciate the regulatory issues and strive to provide solutions that not only advance the science but also address specific challenges. For example, the focus on the critical needs identified herein will serve well both the researchers and the ultimate end-users if targeted collaborations and regulatory-informed case studies to build confidence in their performance and utility. In addition, cross-validation collaborative trials are needed to achieve standardization of the new techniques and analysis methods to build confidence among stakeholders, a relevant example comes from the evolution of the analytical methods used for oil spill response (Faksness et al., 2002).

Concomitantly, the government bodies tasked with enforcing REACH and other relevant regulations, while providing additional useful guidance about the details of chemical characterization of petroleum UVCBs (ECHA, 2022), should have realistic expectations as to what is achievable using even the most advanced analytical methods. The most comprehensive analytical information may or may not be actionable in terms of hazard evaluation in absence of anchoring to the toxicological data. In this regard, existing data on a handful of known hazardous components in petroleum UVCBs, be it “priority PAHs” (ATSDR, 2005) or other constituents with existing human hazard evaluations, are widely regarded as insufficient for assessment of the whole substance(s). Thus, additional data from *in vitro* and other “bioactivity” data streams may be needed to determine the toxicologically relevant compositional features and variability within and among substances and categories (Grimm et al., 2016; House et al., 2022; House et al., 2021).

Concomitantly, the size and complexity of data sets pertaining to sample composition, which are increasingly accompanied with orthogonal high-dimensional information such as (eco)toxicological data, present challenges with respect to data handling and subsequent interpretation. These comprehensive high-/multi-dimensional datasets will typically be subject to dimensionality reduction and used to support grouping and/or classification of the individual petroleum UVCBs. A recent study demonstrated that the choices of the data analysis and visualization methods can not only potentially aid in the communication of “sufficient similarity” among complex substances, but also yield different outcomes in terms of grouping and classification (Onel et al., 2019). The proverbial “black box” of data processing and analysis may create “*a variety of methodological and scientific concerns which mean that it is impossible to independently assess the methods and results*” (ECHA, 2020a) in terms of reliably predicting the properties of the substances that are being evaluated. Hence, it is important to involve not only the analytical chemists when examining the data, but those in other scientific disciplines, such as statistics, mathematics, and

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

computer science, and ensure that the data processing and analysis methods are transparent, and that the decision-makers are sufficiently familiar with the bioinformatics aspects of the information presented to them. Such involvement can expedite assessments of which components are particularly relevant for regulatory needs and may warrant greater focus and analytical precision.

Overall, while statutory and advisory language in the government agencies-produced documents may seem clear, the analytical science may not be available to fully identify and quantify all, or even most, components within complex petroleum UVCBs. It is being acknowledged that "*it is not required to provide detailed structural information on all constituents of a UVCB substance, but there must be sufficient characterization of constituents so as to demonstrate structural similarity, and consecutively provide a basis for predicting the properties of the substance in read-across*" (ECHA, 2020a). Even though additional clarifications have been recently made with respect to what "sufficient" may mean in the context of petroleum UVCBs (ECHA, 2022), ultimately, the flexibility is needed to consider new science and determine if it is "best available" and also fit for a specific decision-context purpose so that the statutes and regulations are ultimately enforceable. While adherence to the most common and established methods is understandable, the industry needs to be more open to the application of the modern analytical chemistry methods to the analysis of the samples and inclusion of such data into regulatory submissions. The traditional approaches that are used for broad characterization of petroleum UVCBs still have an important role to play with respect to substance identification, greater consideration of nontargeted approaches (and how these may also shed new light on emerging hazards), followed by targeted approaches in some cases, is needed. It is also important, however, to recognize that the measurement of the composition of a complex substance is very much influenced by the methods and techniques used; comparisons of data from different laboratories must demonstrate appreciation of such factors if the comparisons are to be meaningful. Therefore, detailed characterization of the technical performance of novel analytical methods is required, in addition to the generation of data for a wide range of samples and applications.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The authors wish to thank Dr. Delina Lyon (Shell and CONCAWE) and Stuart Forbes (CONCAWE) for critical suggestions on this manuscript's text and organization. A.T. Roman-Hubers and A. C. Cordova were supported, in part, by a training grant from the National Institute of Environmental Health Sciences (T32 ES026568).

Funding body information

A.T. Roman-Hubers and A. C. Cordova were supported, in part, by a training grant from the National Institute of Environmental Health Sciences (T32 ES026568).

Abbreviations:

API	American Petroleum Institute

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

ASTM	American Society for Testing and Methods
CAS	Chemical Abstract Service
CCS	collision cross section
DBE	double bond equivalents
ECHA	European Chemical Agency
EINECS	European Inventory of Existing Commercial Chemical Substances
US EPA	United States Environmental Protection Agency
FID	Flame ionization detection
FT-ICR MS	Fourier transform ion cyclotron resonance mass spectrometry
FWHM	full width at half maximum
GC	gas chromatography
GC×GC	two-dimensional gas chromatography
GC-MS	gas chromatography coupled to mass spectrometry
HPLC	high-performance liquid chromatography
IMMS	ion mobility mass spectrometry
IR	infrared
KMD	Kendrick mass defect
LC	liquid chromatography
MS	mass spectrometry
NMR	nuclear magnetic resonance
REACH	Registration, Evaluation, Authorisation and Restriction of Chemicals
TIMS	trapped ion mobility spectrometry
TOF MS	time-of-flight mass spectrometry
UV	ultraviolet
UVCBs	substances of unknown, variable composition, complex reaction products, or biological materials

References

Abib GAP, et al. , 2020. Assessing raw materials as potential adsorbents to remove acidic compounds from Brazilian crude oils by ESI (−) FT-ICR MS. An Acad Bras Cienc. 92, e20200214.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Aeppli C, 2022. Recent advance in understanding photooxidation of hydrocarbons after oil spills. Curr Opin Chem Eng. 36, 100763.

Alam MS, et al. , 2016. Using Variable Ionization Energy Time-of-Flight Mass Spectrometry with Comprehensive GCxGC To Identify Isomeric Species. Anal Chem. 88, 4211–20. [PubMed: 26981641]

Alam MS, et al. , 2018. Mapping and quantifying isomer sets of hydrocarbons ($\geq C_{12}$) in diesel exhaust, lubricating oil and diesel fuel samples using GC × GC-ToF-MS. Atmos Meas Tech. 11, 3047–3058.

Allio L, et al. , 2006. Enhancing the role of science in the decision-making of the European Union. Regul Toxicol Pharmacol. 44, 4–13. [PubMed: 16266777]

API, Petroleum Process Stream Terms Included in the Chemical Substances Inventory Under the Toxic Substances Control Act (TSCA). In: Control, H. S. R. c. T. F. o. T. S., (Ed.). American Petroleum Institute, 1983.

ASTM International, UOP Method 990–11: Organic Analysis of Distillate by Comprehensive Two-Dimensional Gas Chromatography with Flame Ionization Detection. ASTM International, West Conshohocken, PA, 2011.

ATSDR, Toxicology profile for polyaromatic hydrocarbons. CRC Press Boca Raton City, FL, 2005.

Bae E, et al. , 2010. Identification of about 30 000 Chemical Components in Shale Oils by Electrospray Ionization (ESI) and Atmospheric Pressure Photoionization (APPI) Coupled with 15 T Fourier Transform Ion Cyclotron Resonance Mass Spectrometry (FT-ICR MS) and a Comparison to Conventional Oil. Energ Fuel. 24, 2563–2569.

Ball GI, Aluwihare L. l., 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.

Barrow MP, et al. , 2014. An added dimension: GC atmospheric pressure chemical ionization FTICR MS and the Athabasca oil sands. Anal Chem. 86, 8281–8. [PubMed: 25036898]

Barrow MP, et al. , 2016. Effects of Extraction pH on the Fourier Transform Ion Cyclotron Resonance Mass Spectrometry Profiles of Athabasca Oil Sands Process Water. Energy & Fuels. 30, 3615–3621.

Benassi M, et al. , 2013. Petroleum crude oil analysis using low-temperature plasma mass spectrometry. Rapid Commun Mass Spectrom. 27, 825–34. [PubMed: 23495029]

Bierkens J, Geerts L, 2014. Environmental hazard and risk characterisation of petroleum substances: a guided “walking tour” of petroleum hydrocarbons. Environ Int. 66, 182–93. [PubMed: 24607926]

Bowman DT, et al. , 2019. Profiling of individual naphthenic acids at a composite tailings reclamation fen by comprehensive two-dimensional gas chromatography-mass spectrometry. Sci Total Environ. 649, 1522–1531. [PubMed: 30308920]

Bowman DT, et al. , 2020. Isomer-specific monitoring of naphthenic acids at an oil sands pit lake by comprehensive two-dimensional gas chromatography-mass spectrometry. Sci Total Environ. 746, 140985.

Brockman SA, et al. , 2018. Van Krevelen diagram visualization of high resolution-mass spectrometry metabolomics data with OpenVanKrevelen. Metabolomics. 14, 48. [PubMed: 30830359]

Cao XH, et al. , 2020. Sequential thermal dissolution of two low-rank coals and characterization of their structures by high-performance liquid chromatography/time-of-flight mass spectrometry and gas chromatography/mass spectrometry. Rapid Commun in Mass Spectrom. 34, e8887.

Castiblanco JEB, et al. , 2020. Molecular behavior assessment on initial stages of oil spill in terrestrial environments. Environ Sci Pollut Res. 28, 13595–13604.

Chen X, et al. , 2018. Separation and Molecular Characterization of Ketones in a Low-Temperature Coal Tar. Energy Fuels. 32, 4662–4670.

Chen XB, et al. , 2012. Characterization and Comparison of Nitrogen Compounds in Hydrotreated and Untreated Shale Oil by Electrospray Ionization (ESI) Fourier Transform Ion Cyclotron Resonance Mass Spectrometry (FT-ICR MS). Energy & Fuels. 26, 1707–1714.

Cheng X, Hous D, 2021. Characterization of Severely Biodegraded Crude Oils Using Negative-Ion ESI Orbitrap MS, GC-NCD and GC-SCD: Insights into Heteroatomic Compounds Biodegradation. Energies. 14.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Cho Y, et al. , 2017. Extension of the Analytical Window for Characterizing Aromatic Compounds in Oils Using a Comprehensive Suite of High-Resolution Mass Spectrometry Techniques and Double Bond Equivalence versus Carbon Number Plot. Energy & Fuels. 31, 7874–7883.

Cho Y, et al. , 2011. Planar limit-assisted structural interpretation of saturates/aromatics/resins/asphaltenes fractionated crude oil compounds observed by Fourier transform ion cyclotron resonance mass spectrometry. Anal Chem. 83, 6068–73. [PubMed: 21692518]

Cho Y, et al. , 2014. Evaluation of Laser Desorption Ionization Coupled to Fourier Transform Ion Cyclotron Resonance Mass Spectrometry To Study Metalloporphyrin Complexes. Energy & Fuels. 28, 6699–6706.

Clark CR, et al. , 2013. A GHS-consistent approach to health hazard classification of petroleum substances, a class of UVCB substances. Regul Toxicol Pharmacol. 67, 409–20. [PubMed: 24025648]

CONCAWE, REACH – Analytical characterisation of petroleum UVCB substances. Brussels, Belgium, 2012.

CONCAWE, Guidance on Reporting Analytical Information for Petroleum Substances in REACH Registration Dossiers V2. Brussels, 2014.

CONCAWE, Concawe Substance Identification Group Analytical Program Report (Abridged Version). Brussels, Belgium, 2019.

CONCAWE, Guidance to Registrants on Methods for Characterisation of Petroleum UVCB Substances for REACH Registration Purposes. Brussels, Belgium, 2020.

CONCAWE, Petroleum Substances and REACH. 2021.

CONCAWE, REACH Background. 2022.

Damasceno FC, et al. , 2014. Characterization of naphthenic acids using mass spectroscopy and chromatographic techniques: study of technical mixtures. Anal Methods. 6, 807–816.

Dimitrov SD, et al. , 2015. UVCB substances: methodology for structural description and application to fate and hazard assessment. Environ Toxicol Chem. 34, 2450–62. [PubMed: 26053589]

Dodds JN, Baker ES, 2019. Ion Mobility Spectrometry: Fundamental Concepts, Instrumentation, Applications, and the Road Ahead. J Am Soc Mass Spectrom. 30, 2185–2195. [PubMed: 31493234]

Dong X, et al. , 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.

ECHA, Guidance on information requirements and chemical safety assessment. Chapter R.6: QSARS and grouping of chemicals. 2008.

ECHA, Decision on a testing proposal set out in a registration pursuant to Article 40(3) of regulation (EC) No 1907/2006 for Asphalt, CAS No 8052–42-4 (EC No 232–490-9). European Chemical Agency, Helsinki, Finland, 2014.

ECHA, Guidance for identification and naming of substances under REACH and CLP. Vol. 2.1. European Chemical Agency, Helsinki, Finland, 2017a.

ECHA, Guidance on requirements for substances in articles. Vol. Version 4.0. European Chemicals Agency, Helsinki, Finland, 2017b.

ECHA, Read-Across Assessment Framework (RAAF) - considerations on multi-constituent substances and UVCBs. European Chemical Agency, Helsinki, Finland, 2017c.

ECHA, Testing Proposal Decision on Substance EC 295–332-8 “Extracts (petroleum), deasphalted vacuum residue solvent”. European Chemicals Agency, Helsinki, Finland, 2020a.

ECHA, Testing Proposals Decision on Substance EC 265–110-5 “Extracts (petroleum), residual oil solvent”. 2020b.

ECHA, Testing Proposal Decision on Substance EC 265–182-8 “Gas oils (petroleum), hydrodesulfurized”. European Chemicals Agency, Helsinki, Finland, 2021.

ECHA, Advice on using read-across for UVCB substances - obligations arising from Commission Regulation 2021/979, amending REACH annexes. European Chemicals Agency, Helsinki, Finland, 2022.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

EPA, U. S., Toxic Substances Control Act Inventory Representation for Chemical Substances of Unknown or Variable Composition, Complex Reaction Products and Biological Materials: UVCB Substances. 1995.

European Commission, 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). Official Journal of the European Union, Brussels, Belgium, 2021.

European Council, Corrigendum to Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC. Vol. L 136/3. European Commission, Brussels, Belgium, 2007.

Faksness LG, et al. , 2002. Round Robin Study—Oil Spill Identification. Environ Forensics. 3, 279–291.

Fernandez-Lima FA, et al. , 2009. Petroleum crude oil characterization by IMS-MS and FTICR MS. Anal Chem. 81, 9941–7. [PubMed: 19904990]

Folkerts EJ, et al. , 2019. Toxicity in aquatic model species exposed to a temporal series of three different flowback and produced water samples collected from a horizontal hydraulically fractured well. Ecotoxicol Environ Saf. 180, 600–609. [PubMed: 31132555]

Frenzel M, et al. , 2010. Complications with remediation strategies involving the biodegradation and detoxification of recalcitrant contaminant aromatic hydrocarbons. Sci Total Environ. 408, 4093–4101. [PubMed: 20542318]

FuelsEurope, FuelsEurope position paper on REACH and the Refining industry. Brussels, Belgium, 2015.

Gabetti E, et al. , 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.

Gavard R, et al. , 2020. KairosMS: A New Solution for the Processing of Hyphenated Ultrahigh Resolution Mass Spectrometry Data. Anal Chem. 92, 3775–3786. [PubMed: 31990191]

Giles HN, 2016. Crude Oil Analysis: History and Development of Test Methods From 1854 to 2016. Mater Perform Charact. 5, 1–169.

Gough MA, Rowland SJ, 1990. Characterization of unresolved complex mixtures of hydrocarbons in petroleum. Nature. 344, 648–650.

Grimm FA, et al. , 2016. A chemical-biological similarity-based grouping of complex substances as a prototype approach for evaluating chemical alternatives. Green Chem. 18, 4407–4419. [PubMed: 28035192]

Haitao S, et al. , 2013. Hydrocarbon composition of different VGO feedstocks and the correlation with FCC product distribution. China Petrol Proc Petrochem Techn. 15.

Hao J, et al. , 2017. Thermal cracking behaviors and products distribution of oil sand bitumen by TG-FTIR and Py-GC/TOF-MS. Energy Convers Manag. 151, 227–239.

Headley JV, et al. , 2011. Preliminary fingerprinting of Athabasca oil sands polar organics in environmental samples using electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. Rapid Commun Mass Spectrom. 25, 1899–909. [PubMed: 21638366]

Hoskins JN, et al. , 2011. Architectural Differentiation of Linear and Cyclic Polymeric Isomers by Ion Mobility Spectrometry-Mass Spectrometry. Macromolecules. 44, 6915–6918.

Hosseini SH, et al. , 2021. Characterization of crude oils derived from carbonate and siliciclastic source rocks using FTICR-MS. Org Geochem. 159, 104286.

Hourani N, et al. , 2013. Atmospheric pressure chemical ionization Fourier transform ion cyclotron resonance mass spectrometry for complex thiophenic mixture analysis. Rapid Commun Mass Spectrom. 27, 2432–8. [PubMed: 24097400]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

House JS, et al. , 2022. Grouping of UVCB substances with dose-response transcriptomics data from human cell-based assays. ALTEX. 39, 388–404. [PubMed: 35288757]

House JS, et al. , 2021. Grouping of UVCB substances with new approach methodologies (NAMs) data. ALTEX. 38, 123–137. [PubMed: 33086383]

Hsu CS, et al. , 2011. Petroleomics: advanced molecular probe for petroleum heavy ends. J Mass Spectrom. 46, 337–43. [PubMed: 21438082]

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

Hu M, et al. , 2018. Collision cross section (CCS) measurement by ion cyclotron resonance mass spectrometry with short-time Fourier transform. Rapid Commun Mass Spectrom. 32, 751–761. [PubMed: 29486526]

Hughey CA, et al. , 2001. Kendrick mass defect spectrum: a compact visual analysis for ultrahigh-resolution broadband mass spectra. Anal Chem. 73, 4676–81. [PubMed: 11605846]

Hughey CA, et al. , 2002. Resolution of 11,000 compositionally distinct components in a single electrospray ionization Fourier transform ion cyclotron resonance mass spectrum of crude oil. Anal Chem. 74, 4145–9. [PubMed: 12199586]

Jaggi A, et al. , 2019. Composition of the dissolved organic matter produced during in situ burning of spilled oil. Org Geochem. 138, 103926.

Jennerwein MK, et al. , 2014. Complete Group-Type Quantification of Petroleum Middle Distillates Based on Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry (GC×GC-TOFMS) and Visual Basic Scripting. Energy Fuels. 28, 5670–5681.

Kaiser MJ, 2017. A review of refinery complexity applications. Pet Sci. 14, 167–194.

Kendrick E, 1963. A Mass Scale Based on CH2 = 14.0000 for High Resolution Mass Spectrometry of Organic Compounds. Anal. Chem 35, 2146–2154.

Kew W, et al. , 2017. Interactive van Krevelen diagrams - Advanced visualisation of mass spectrometry data of complex mixtures. Rapid Commun Mass Spectrom. 31, 658–662. [PubMed: 28063248]

Kim D, et al. , 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 S, et al. , 2003. Graphical method for analysis of ultrahigh-resolution broadband mass spectra of natural organic matter, the van Krevelen diagram. Anal Chem. 75, 5336–44. [PubMed: 14710810]

Kondyli A, Schrader W, 2019. High-resolution GC/MS studies of a light crude oil fraction. J Mass Spectrom. 54, 47–54. [PubMed: 30378212]

Kulkarni SU, Thies MC, 2012. Quantitative analysis of polydisperse systems via solvent-free matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Rapid Commun Mass Spectrom. 26, 392–398. [PubMed: 22223328]

Lacroix-Andrivet O, et al. , 2021. Molecular Characterization of Aged Bitumen with Selective and Nonselective Ionization Methods by Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. 2. Statistical Approach on Multiple-Origin Samples. Energy & Fuels. 35, 16442–16451.

Lai A, et al. , 2022. The Next Frontier of Environmental Unknowns: Substances of Unknown or Variable Composition, Complex Reaction Products, or Biological Materials (UVCBs). Environ Sci Technol. 56, 7448–7466. [PubMed: 35533312]

Lalli PM, et al. , 2015. Isomeric Separation and Structural Characterization of Acids in Petroleum by Ion Mobility Mass Spectrometry. Energy Fuels. 29, 3626–3633.

Lalli PM, et al. , 2017. Functional Isomers in Petroleum Emulsion Interfacial Material Revealed by Ion Mobility Mass Spectrometry and Collision-Induced Dissociation. Energy & Fuels. 31, 311–318.

Leefmann T, et al. , 2019. UltraMassExplorer: a browser-based application for the evaluation of high-resolution mass spectrometric data. Rapid Commun Mass Spectrom. 33, 193–202. [PubMed: 30366355]

Li S, et al. , 2022a. Molecular characteristics of biochar-derived organic matter sub-fractions extracted by ultrasonication. Sci Total Environ. 806, 150190.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Li Y, et al. , 2022b. Comprehensive chemical characterization of dissolved organic matter in typical point-source refinery wastewaters. Chemosphere. 286, 131617.

Liu D, et al. , 2014. Direct hydro-liquefaction of sawdust in petroleum ether and comprehensive bio-oil products analysis. Bioresour Technol. 155, 152–60. [PubMed: 24445192]

Liu Y, et al. , 2020. The acid and neutral nitrogen compounds characterized by negative ESI Orbitrap MS in a heavy oil before and after oxidation. Fuel. 277.

Lu H, et al. , 2013. Geochemical Explication of Sulfur Organics Characterized by Fourier Transform Ion Cyclotron Resonance Mass Spectrometry on Sulfur-Rich Heavy Oils in Jinxian Sag, Bohai Bay Basin, Northern China. Energy Fuels. 27, 5861–5866.

Luna N, et al. , 2014a. Identification And Characterization Of Sulfur Compounds In Straight-Run Diesel Using Comprehensive Two-Dimensional Gas Chromatography Coupled To Time-Of-Flight Mass Spectrometry. China Pet Process Petrochem. 16, 10–18.

Luna N, et al. , 2014b. Identification and Characterization of Sulfur Compounds in Straight-Run Diesel Using Comprehensive Two-Dimensional GC Coupled with TOF MS. China Petrol Proc Petrochem Techn. 16, 10–18.

Luo T, et al. , 2016. A novel characterization of furfural-extract oil from vacuum gas oil and its application in solvent extraction process. Fuel Process Technol. 152, 356–366.

Lv H, et al. , 2013. Application of UPLC-Quadrupole-TOF-MS Coupled with Recycling Preparative HPLC in Isolation and Preparation of Coumarin Isomers with Similar Polarity from Peucedanum praeruptorum. Chromatographia. 76, 141–148.

MacLennan MS, et al. , 2018. Characterization of Athabasca lean oil sands and mixed surficial materials: Comparison of capillary electrophoresis/low-resolution mass spectrometry and high-resolution mass spectrometry. Rapid Commun Mass Spectrom. 32, 695–702. [PubMed: 29486520]

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. [PubMed: 30172356]

Maillard JF, et al. , 2021. Structural analysis of petroporphyrins from asphaltene by trapped ion mobility coupled with Fourier transform ion cyclotron resonance mass spectrometry. Analyst. 146, 4161–4171. [PubMed: 34047731]

Mao D, et al. , 2009. Combining HPLC-GCXGC, GCXGC/ToF-MS, and selected ecotoxicity assays for detailed monitoring of petroleum hydrocarbon degradation in soil and leaching water. Environ Sci Technol. 43, 7651–7. [PubMed: 19921874]

Marshall AG, Rodgers RP, 2004. Petroleomics: the next grand challenge for chemical analysis. Acc Chem Res. 37, 53–9. [PubMed: 14730994]

Marshall AG, Rodgers RP, 2008. Petroleomics: chemistry of the underworld. Proc Natl Acad Sci U S A. 105, 18090–5. [PubMed: 18836082]

McKee RH, et al. , 2015. Characterization of the toxicological hazards of hydrocarbon solvents. Crit Rev Toxicol. 45, 273–365. [PubMed: 25868376]

McKenna AM, et al. , 2013. Expansion of the analytical window for oil spill characterization by ultrahigh resolution mass spectrometry: beyond gas chromatography. Environ Sci Technol. 47, 7530–9. [PubMed: 23692145]

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, 7348–7353.

Miles SM, et al. , 2020. Oil sands process affected water sourced Trichoderma harzianum demonstrates capacity for mycoremediation of naphthenic acid fraction compounds. Chemosphere. 258, 127281.

Muller H, et al. , 2020. Innate Sulfur Compounds as an Internal Standard for Determining Vacuum Gas Oil Compositions by APPI FT-ICR MS. Energy Fuels. 34, 8260–8273.

National Research Council, 2009. Science and Decisions: Advancing Risk Assessment. National Academies Press, Washington, DC.

Neumann A, et al. , 2021. Investigation of Island/Single-Core- and Archipelago/Multicore-Enriched Asphaltenes and Their Solubility Fractions by Thermal Analysis Coupled with High-Resolution

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. Energy & Fuels. 35, 3808–3824.

Ngo HI, et al. , 2012. Improved synthesis and characterization of saturated branched-chain fatty acid isomers. Eur J Lipid Sci Technol. 114, 213–221.

Niyonsaba E, et al. , 2019. Recent Advances in Petroleum Analysis by Mass Spectrometry. Anal Chem. 91, 156–177. [PubMed: 30428670]

Nolte T, et al. , 2013. Desulfurized Fuels from Athabasca Bitumen and Their Polycyclic Aromatic Sulfur Heterocycles. Analysis Based on Capillary Electrophoresis Coupled with TOF MS. Energy Fuels. 27, 97–107.

O'Reilly KT, et al. , 2019. Oxygen-Containing Compounds Identified in Groundwater from Fuel Release Sites Using GCxGC-TOF-MS. Ground Water Monit R. 39, 32–40.

Oldenburg TBP, et al. , 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, et al. , 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, et al. , 2019. Grouping of complex substances using analytical chemistry data: A framework for quantitative evaluation and visualization. PLoS One. 14, e0223517.

Orrego- Ruíz JA, 2018. Finding a relationship between the composition and the emulsifying character of asphaltenes through FTICR-MS. Cienc Tecn Fut. 8, 45–52.

Palacio Lozano DC, et al., 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, et al. , 2019b. Pushing the analytical limits: new insights into complex mixtures using mass spectra segments of constant ultrahigh resolving power. Chem Sci. 10, 6966–6978. [PubMed: 31588263]

Palacio Lozano DC, et al. , 2022. Revealing the Reactivity of Individual Chemical Entities in Complex Mixtures: the Chemistry Behind Bio-Oil Upgrading. Anal Chem. 94, 7536–7544. [PubMed: 35576165]

Palacio Lozano DC, et al. , 2020. Petroleomics: Tools, Challenges, and Developments. Annu Rev Anal Chem (Palo Alto Calif). 13, 405–430. [PubMed: 32197051]

Pereira AS, et al. , 2013. Characterization of oil sands process-affected waters by liquid chromatography orbitrap mass spectrometry. Environ Sci Technol. 47, 5504–13. [PubMed: 23607765]

Petroleum HPV Testing Group, HPV Challenge Overview. API, 2017.

Phillips AL, et al. , 2022. A Framework for Utilizing High-Resolution Mass Spectrometry and Nontargeted Analysis in Rapid Response and Emergency Situations. Environ Toxicol Chem. 41, 1117–1130. [PubMed: 34416028]

Porto CFC, et al. , 2019. Characterization of organosulfur compounds in asphalt cement samples by ESI(+)FT-ICR MS and 13C NMR spectroscopy. Fuel. 256, 115923.

Prince RC, Walters CC, 2022. Modern analytical techniques are improving our ability to follow the fate of spilled oil in the environment. Curr Opin Chem Eng. 36, 100787.

Qian K, et al. , 2004. The Coupling of Supercritical Fluid Chromatography and Field Ionization Time-of-Flight High-Resolution Mass Spectrometry for Rapid and Quantitative Analysis of Petroleum Middle Distillates. Eur J Mass Spectrom. 10, 187–196.

Qian K, et al. , 2001. Reading chemical fine print: Resolution and identification of 3000 nitrogen-containing aromatic compounds from a single electrospray ionization Fourier transform ion cyclotron resonance mass spectrum of heavy petroleum crude oil. Energy & Fuels. 15, 492–498.

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, 2785–2794. [PubMed: 31741268]

Rasmussen K, et al. , 1999. Compilation of EINECS: Descriptions and definitions used for UVCB substances: Complex reaction products, plant products, (post-reacted) naturally occurring

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

substances, micro-organisms, petroleum products, soaps and detergents, and metallic compounds. Toxicol Environ Chem. 69, 403–416.

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. Mar Pollut Bull. 38, 126–135.

Redman AD, et al. , 2014. PETRORISK: a risk assessment framework for petroleum substances. Integr Environ Assess Manag. 10, 437–48. [PubMed: 24687890]

Redman AD, et al. , 2012. PETROTOX: an aquatic toxicity model for petroleum substances. Environ Toxicol Chem. 31, 2498–506. [PubMed: 22887448]

Riches E, et al., Ion Mobility & PetroOrg Software: Novel Techniques for Petroleomics Investigations In: Waters, (Ed.), 2015.

Ristic ND, et al. , 2018. Compositional Characterization of Pyrolysis Fuel Oil from Naphtha and Vacuum Gas Oil. Energy Fuels. 32, 1276–1286.

Rocha YDS, et al. , 2018. Geochemical characterization of lacustrine and marine oils from offshore 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, 4665–87. [PubMed: 21528862]

Rodgers RP, et al. , 1998. Resolution, elemental composition, and simultaneous monitoring by Fourier transform ion cyclotron resonance mass spectrometry of organosulfur species before and after diesel fuel processing. Anal Chem. 70, 4743–4750.

Rodrigues Covas T, et al. , 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, 117289.

Roman-Hubers AT, et al. , 2021. Data Processing Workflow to Identify Structurally Related Compounds in Petroleum Substances Using Ion Mobility Spectrometry-Mass Spectrometry. Energy Fuels. 35, 10529–10539. [PubMed: 34366560]

Roman-Hubers AT, et al. , 2022. Characterization of Compositional Variability in Petroleum Substances. Fuel. 317, 123547.

Rosa TR, et al. , 2020. DropMS: Petroleomics Data Treatment Based in Web Server for High-Resolution Mass Spectrometry. J Am Soc Mass Spectrom. 31, 1483–1490. [PubMed: 32519862]

Rowland SJ, et al. , 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. [PubMed: 21598334]

Ruger CP, et al. , 2015. Hyphenation of Thermal Analysis to Ultrahigh-Resolution Mass Spectrometry (Fourier Transform Ion Cyclotron Resonance Mass Spectrometry) Using Atmospheric Pressure Chemical Ionization For Studying Composition and Thermal Degradation of Complex Materials. Anal Chem. 87, 6493–9. [PubMed: 26024433]

Rui D, et al. , 2012. Molecular Characterization of Hydrotreated Atmospheric Residue Derived from Arabian Heavy Crude by GC FI/FD TOF MS and APPI FT-ICR MS. China Pet Process Petrochem. 14, 80–88.

Salvito D, 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, 2097–2108. [PubMed: 32780492]

Santos JM, et al. , 2015. Petroleomics by ion mobility mass spectrometry: resolution and characterization of contaminants and additives in crude oils and petrofuels. Anal Methods. 7, 4450–4463.

Scarlett A, et al. , 2008. Chronic sublethal effects associated with branched alkylbenzenes bioaccumulated by mussels. Environ Toxicol Chem. 27, 561–567. [PubMed: 17988179]

Sierra Analytics, Composer: State of the art visualization software for petrolemics. 2022.

Silva RC, et al. , 2020. Mechanistic insights into sulfur rich oil formation, relevant to geological carbon storage routes. A study using (+) APPI FTICR-MS analysis. Org Geochem. 147.

Silva RVS, et al. , 2019. Comprehensive study of the liquid products from slow pyrolysis of crambe seeds: Bio-oil and organic compounds of the aqueous phase. Biomass Bioenerg. 123, 78–88.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Smith HM, et al. , 1959. Keys to the mystery of crude oil. Proc. Amer. Petrole. Inst 39, 433–465.

Sorensen L, et al. , 2019. Establishing a link between composition and toxicity of offshore produced waters using comprehensive analysis techniques - A way forward for discharge monitoring? Sci Total Environ. 694, 133682.

Spectroswiss, FTMS Data Processing Tools PeakByPeak FTMS Data Analysis. In: Spectroswiss, (Ed.), 2019.

Stenson AC, et al. , 2003. Exact masses and chemical formulas of individual Suwannee River fulvic acids from ultrahigh resolution electrospray ionization Fourier transform ion cyclotron resonance mass spectra. Anal Chem. 75, 1275–84. [PubMed: 12659186]

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, (Eds.), Oil Spill Environmental Forensics: Fingerprinting And Source Identification. Academic Press, Cambridge, MA, pp. 1–53.

U.S. EPA, Toxic Substance Control Act (TSCA) PL 94–469 Candidate List of Chemical Substances Addendum 1 Generic Terms Covering Petroleum Refinery Process Streams. US Environmental Protection Agency,, Washigton, D.C., 1978.

Urban, P. L., 2016. Quantitative mass spectrometry: an overview. Philos Trans A Math Phys Eng Sci. 374, 20150382.

US EPA, Method 8051B: Non-halogenated organics using GC-FID. US Environmental Protection Agency, Washington, DC, 1996.

US EPA, Method 8270E (SW-846): Semivolatile Organic Compounds by Gas Chromatography/Mass Spectrometry (GC/MS). US Environmental Protection Agency, Washington, DC, 2014.

US EPA, Procedures for Chemical Risk Evaluation Under the Amended Toxic Substances Control Act. 82 FR 33726, Vol. 40 CFR 702. US Environmental Protection Agency, Washington, DC, 2017.

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, et al. , 2020. Characterization of nonvolatile polar compounds from Brazilian oils by electrospray ionization with FT-ICR MS and Orbitrap-MS. Fuel. 282, 118790.

Ventura GT, et al. , 2011. Analysis of petroleum compositional similarity using multiway principal components analysis (MPCA) with comprehensive two-dimensional gas chromatographic data. J Chromatogr A. 1218, 2584–92. [PubMed: 21450298]

Vozka P, et al. , 2019. Jet fuel density via GC x GC-FID. Fuel. 235, 1052–1060.

Walters CC, et al. , 2015. Petroleum alteration by thermochemical sulfate reduction – A comprehensive molecular study of aromatic hydrocarbons and polar compounds. Geochim Cosmochim Acta. 153, 37–71.

Wang X, et al. , 2020. Characterization of wastewater effluent organic matter with different solid phase extraction sorbents. Chemosphere. 257, 127235.

Wang Z, et al. , 1994. Fractionation of a Light Crude Oil and Identification and Quantitation of Aliphatic, Aromatic, and Biomarker Compounds by GC-FID and GC-MS, Part I. J Chromatogr Sci. 32, 361–366.

Wang Z, et al. , 2011. Forensic fingerprinting and source identification of the 2009 Sarnia (Ontario) oil spill. J Environ Monit. 13, 3004–17. [PubMed: 21956546]

Wang ZD, Fingas M, 2003. Fate and identification of spilled oils and petroleum products in the environment by GC-MS and GC-FID. Energy Sources. 25, 491–508.

Weng N, et al. , 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. [PubMed: 25939738]

Wise SA, et al. , 2022. Advances in Chemical Analysis of Oil Spills Since the Deepwater Horizon Disaster. Crit Rev Anal Chem. 1–60.

Wozniak AS, et al. , 2019. Rapid Degradation of Oil in Mesocosm Simulations of Marine Oil Snow Events. Environ Sci Technol. 53, 3441–3450. [PubMed: 30827089]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Xia Y, et al. , 2021. Characterization of nitrogen-containing compounds in petroleum fractions by online reversed-phase liquid chromatography-electrospray ionization Orbitrap mass spectrometry. Fuel. 284, 119035.

Xian F, et al. , 2012. High resolution mass spectrometry. Anal Chem. 84, 708–19. [PubMed: 22263633]

Yang L, et al. , 2019. Gas chromatography-Orbitrap mass spectrometry screening of organic chemicals in fly ash samples from industrial sources and implications for understanding the formation mechanisms of unintentional persistent organic pollutants. Sci Total Environ. 664, 107–115. [PubMed: 30739845]

Zhong JY, et al. , 2011. Combining advanced NMR techniques with ultrahigh resolution mass spectrometry: A new strategy for molecular scale characterization of macromolecular components of soil and sedimentary organic matter. Org Geochem. 42, 903–916.

Zhu GY, et al. , 2020. Discovery and Molecular Characterization of Organic Caged Compounds and Polysulfanes in Zhongba81 Crude Oil, Sichuan Basin, China. Energy Fuels. 34, 6811–6821.

Author Manuscript

Highlights

- Petroleum UVCBs are compositionally complex and variable
- Analytical characterization of these substance is a challenge for registration and evaluation
- Recently, the regulatory authorities provided more specificity about UVCB characterization
- Science advances and novel analytical methods in petroleomics offer solutions for data gaps
- This review aims to bridge the divide between the science of petroleomics and the regulators

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

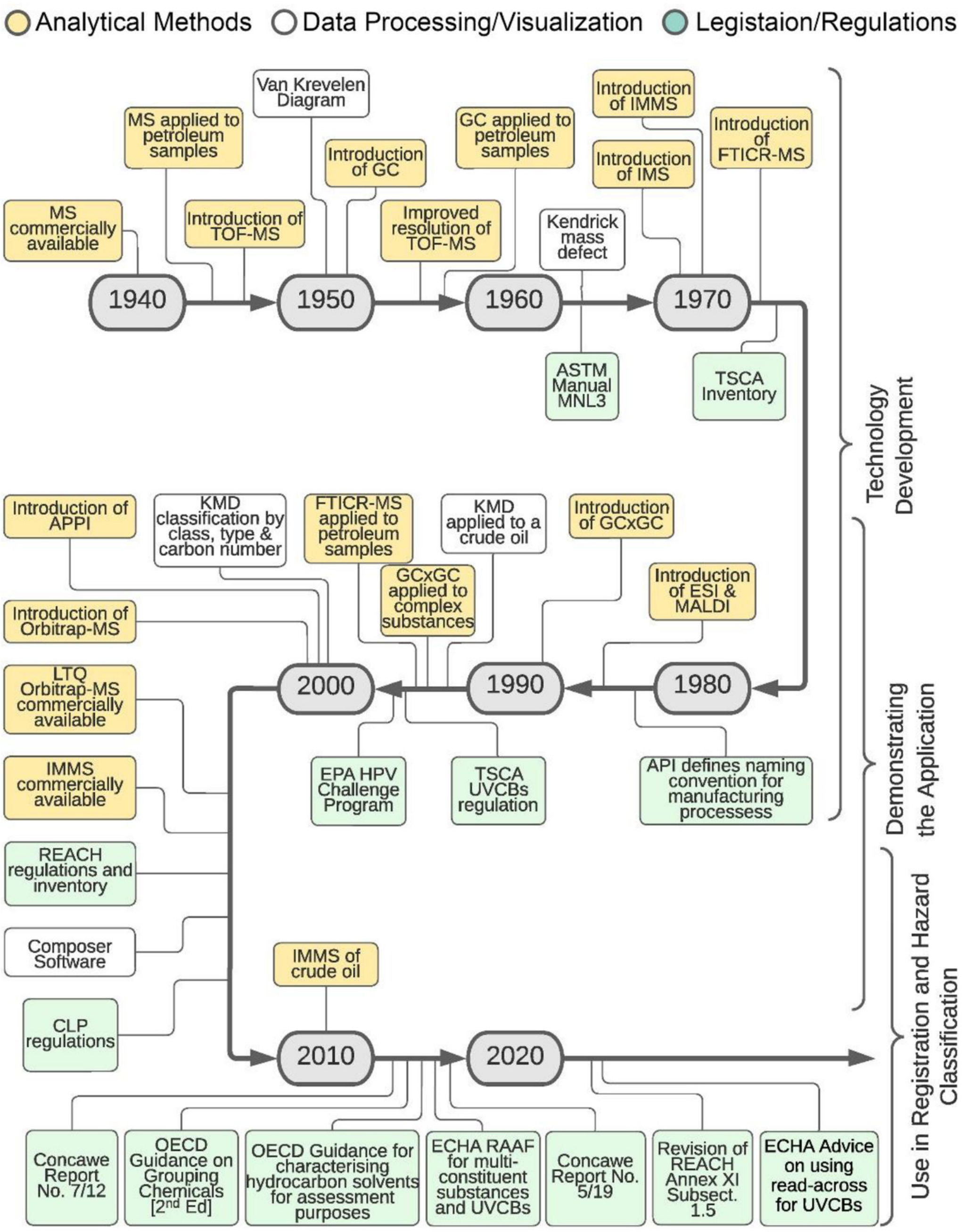

Figure 1.

A timeline of major developments in the fields of analytical chemistry and data analysis of petroleum UVCB, and the concomitant evolution of the regulatory frameworks for registration and hazard classification of these substances See abbreviations in the text.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

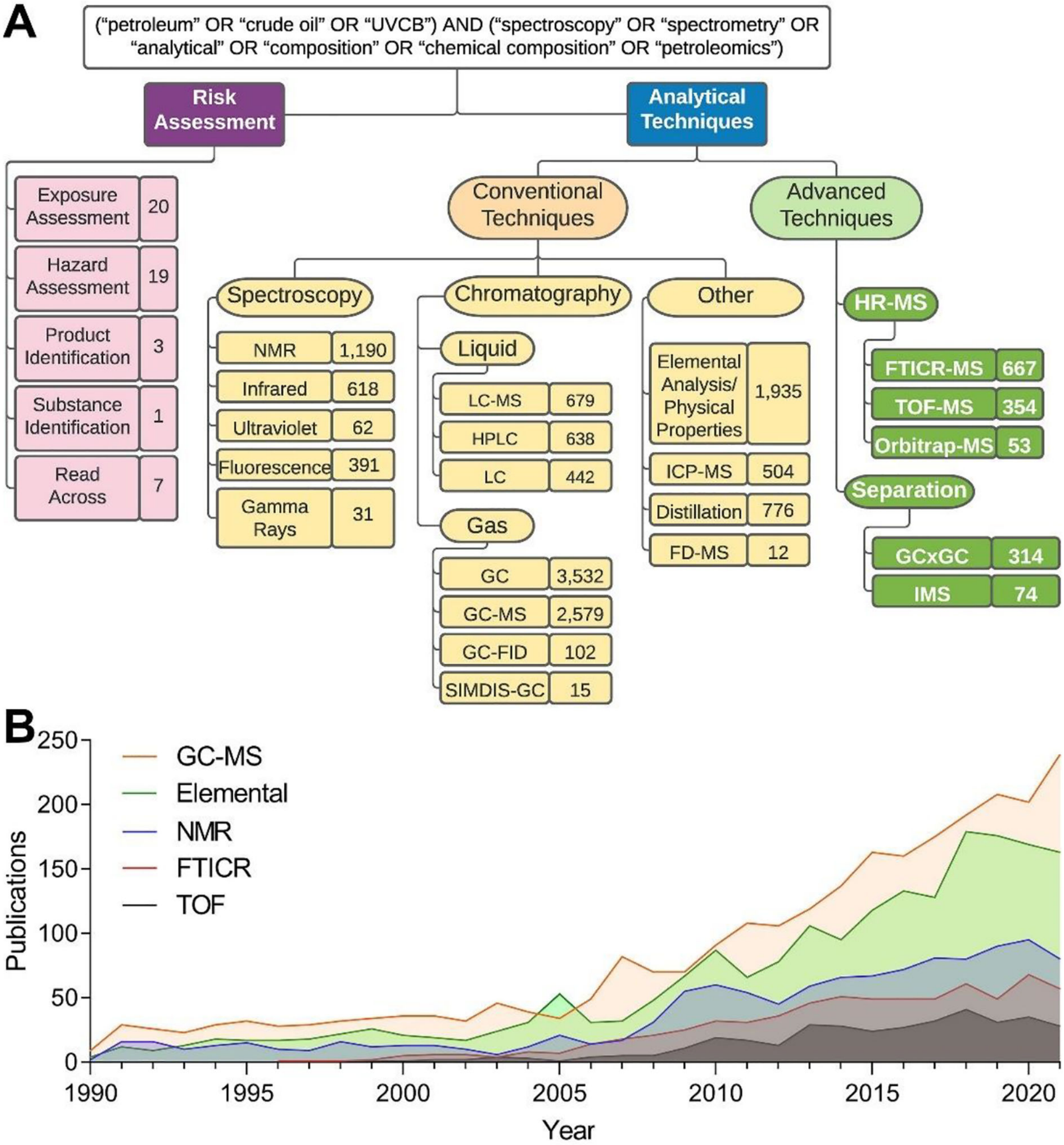

Figure 2.

Literature review of the major analytical methods and their application for the analysis of petroleum substances. **(A)** A dendrogram of the major searches. The numbers indicate the quantity of publications for each search. See Supplemental Table 1 for information on the exact search terms and hyperlinks to the publications. **(B)** Cumulative histograms indicating the number of publications across time through December 2021. Colors correspond to the methods indicated in the inset.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

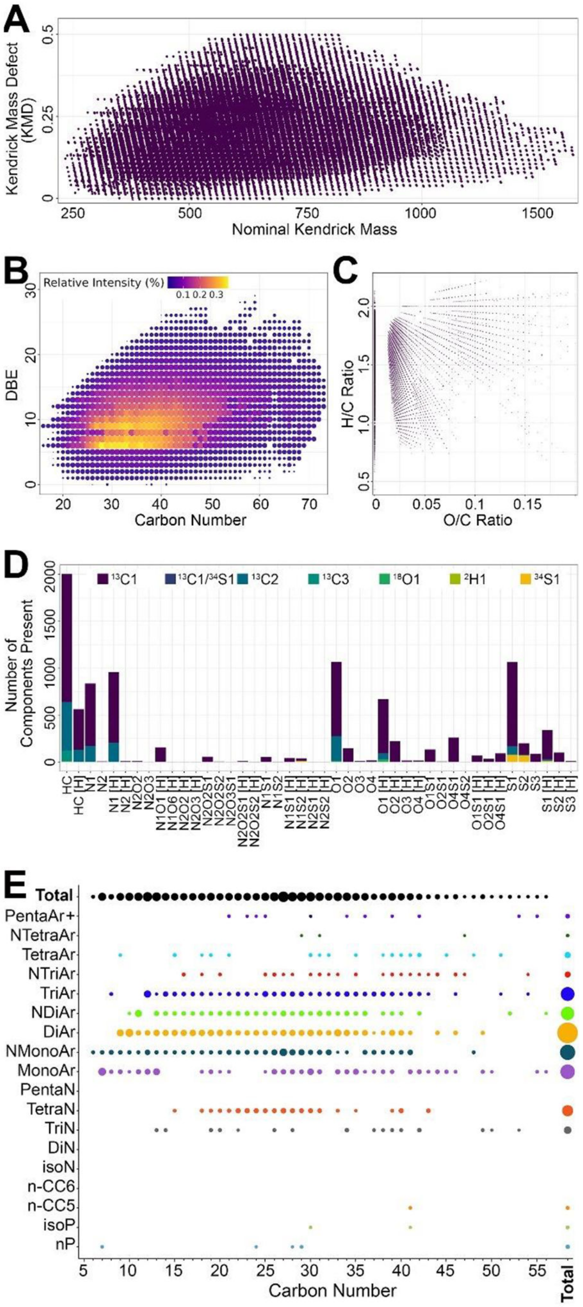

Figure 3.

Example visualizations commonly used to represent hydrocarbon composition of petroleum substances analyzed using high resolution MS techniques. **(A)** A Kendrick mass defect (KMD) plot demonstrating repetitive patterns of CH_2-containing molecules in a petroleum sample. **(B)** A DBE vs carbon number plot indicating the relative proportions of molecules varying by their degree of aromaticity and carbon number in a sample. **(C)** van Krevelen plot display the degree of oxidation by plotting the H/C versus O/C ratio in a sample. **(D)** A stacked bar plot showing relative proportions of constituents from various chemical groups. **(E)** A plot of relative amounts of various hydrocarbon blocks in a sample.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

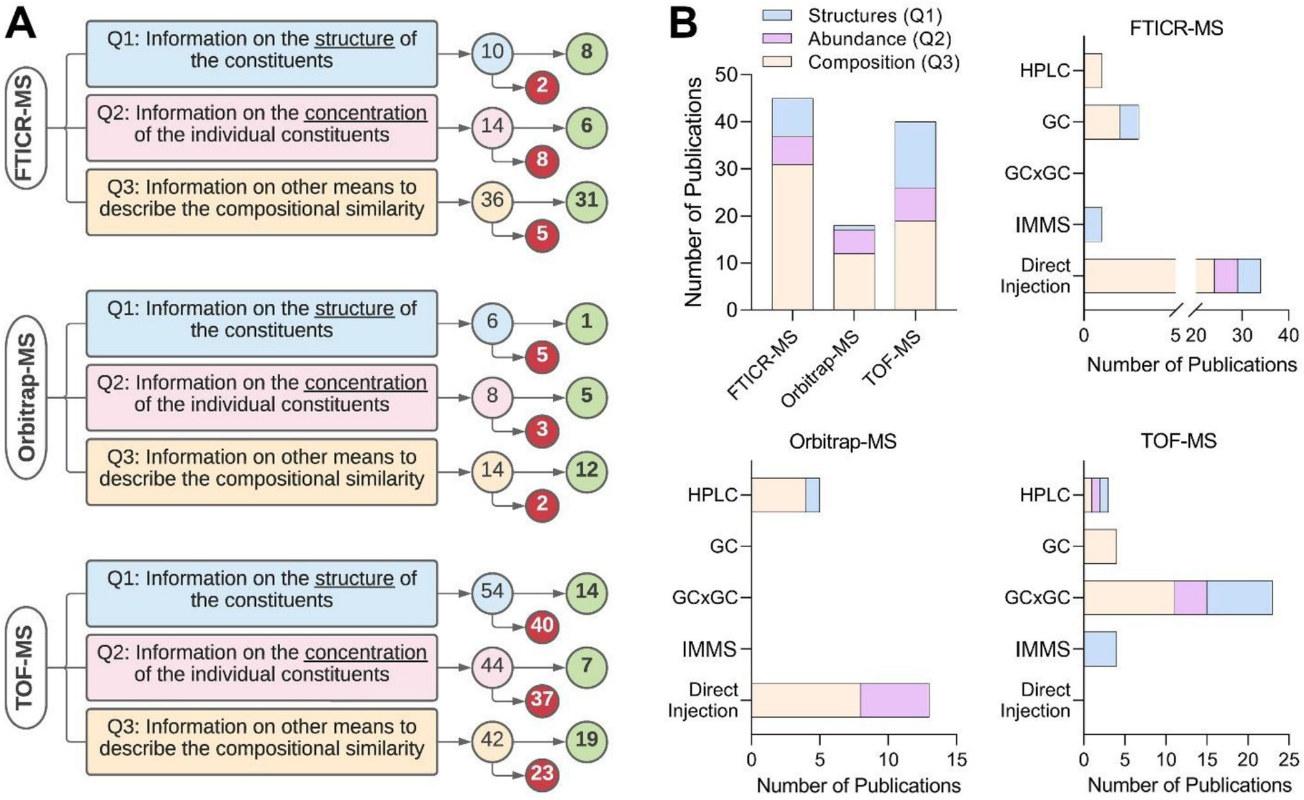

Figure 4.

The use of various novel MS methods to address specific regulatory needs identified in this review. **(A)** Scholarly publications that were identified as relevant to each regulatory need/question (identified by colors). A total number of publications identified by a literature search is listed in the first circle (see Supplemental Table 2 for details). Upon examination of each study's content, a number of publications were deemed not relevant (red circles); the remaining studies are shown in green circles. **(B)** The number of relevant publications as a function of the high-/ultra-high MS technique. Top left, a stacked bar graph indicating the number of publications as they pertain to each regulatory need/question in (A). Remaining stacked bar plots show the number of studies that used various separation (HPLC, GC, GC×GC or IMMS) or direct injection with each MS technique.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 1.

Summary of the chemical characterization needs for UVCB that are registered and evaluated under REACH in the European Union.

Information Needed	Substance Identification for Registration (REACH Annex VI, Section 2)	Hazard Characterization for Evaluation (REACH Annex XI, Section 1.5)		
		Read-Across Assessment Framework: Considerations on multi-constituent substances and UVCBs (ECHA, 2017c)		Advice on Using Read-Across for UVCB substances (ECHA, 2022)
		Category Approach	**Analogue Approach**	
Overall Substance Identification	• Substance name [from a trade association and/or nomenclature system] • Substance identifiers [source, manufacturing process, carbon and boiling range, phys-chem properties, etc.]	• Substance(s) to be grouped in a category • Compositions to be included • Manufacturing process description	• Source substance(s) and target substance • Compositions to be included • Manufacturing process description	• Similarity may be established based on: (i) presence of *identical* constituents OR (ii) variation in concentration and variability in constituents • Constituents present at >1% must be identified; lower thresholds if constituents of concern are present (>0.1% for carcinogenic/ mutagenic, >0.3% for repro/ developmental) • >80% of constituents in the substance must be identified
Composition Characterization	• Constituents present >10% • Constituents <10% that may be impacting hazard classification	Category “domain” needs to be defined (constituent-specific concentration determination not specified)	• Source substance(s) • Target substance (constituent-specific concentration determination not specified)	• Must be characterized up to 100% • When full characterization is impractical/ impossible need to provide (i) justification AND (ii) demonstration of similarity “by other means”
Variability	Not Required	• Structural similarity for category based on worst-case scenario • Determine if quantitative differences or patterns in	• Structural Similarity between source & target based on worst-case scenario • Determine if quantitative differences or	• Structural similarity explained based on quantitative and qualitative comparison of composition • When full characterization

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Information Needed	**Substance Identification for Registration (REACH Annex VI, Section 2)**	**Hazard Characterization for Evaluation (REACH Annex XI, Section 1.5)**		
		Read-Across Assessment Framework: Considerations on multi-constituent substances and UVCBs (ECHA, 2017c)		**Advice on Using Read-Across for UVCB substances (ECHA, 2022)**
		Category Approach	**Analogue Approach**	
		predicted properties may be reflected in structural similarity	patterns in predicted properties may be reflected in structural similarity	is impractical/ impossible need to provide (i) comparison of constituents AND (ii) demonstration of similarity “by other means” (e.g. analytical information for >95% constituents, constituents of high concern, high resolution for confidence fingerprinting) • Analysis of at least 5 independent (i.e., production batches) samples analyzed from ALL registrants of a substance