Extended characterization of petroleum aromatics using off-line LC-GC-MS

Khoa Huynh, Annette Eva Jensen, Jonas Sundberg[1*]

Danish Hydrocarbon Research and Technology Centre, Technical University of Denmark

Elektrovej 375, 2800 Kgs. Lyngby, Denmark

* Corresponding author

E-mail jonsun@dtu.dk ; Phone +45-53 83 19 42

Abstract

Crude oil remains a grand challenge for analytical chemists. With the advent of multi-dimensional chromatography and high-resolution mass spectrometry, an impressive number of compounds have been identified. However, the large diversity in structure and abundance makes it difficult to obtain full compound coverage in a single analysis. Sample preparation methods such as solid-phase extraction and SARA-type analysis reduces this complexity. However, the molecular diversity within each fraction is still highly complex. Thus, in the routine analysis, only a small part of the chemical space is typically characterized. Obtaining a more detailed composition of crude oil is important for production, processing and environmental aspects. We have developed a high-resolution fractionation method for isolation and preconcentration of trace aromatics, including oxygenated and nitrogen-containing species. By the isolation of the more abundant aromatics, i.e. monoaromatics and naphthalenes, trace species can be enriched for analysis. This enables the identification of features not detectable by routine methods. We demonstrate the applicability by fractionation and subsequent GC-MS analysis of 14 crude oils sourced from the North Sea. The number of tentatively identified compounds increased by approximately 60 to 150% compared to solid-phase extraction and GC×GC-MS. Furthermore, the method was used to successfully isolate and identify a new set of heteroatom-containing aromatics (amines, ketones). The method is not intended to replace traditional sample preparation techniques or multi-dimensional chromatography but acts as a complementary tool. An in-depth comparison to routine characterization techniques is presented concerning advantages and disadvantages.

1 Introduction

The use of petroleum as a feedstock for energy production is declining. However, certain critical functions cannot safely be replaced by renewable energy.[1] Secondly. petroleum is a fundamental feedstock for the production of a large number of chemical starting materials.[2,3] Therefore, reducing the environmental impact of oil production is an important goal. This requires a better understanding of petroleum on the molecular level. Crude oil is a complex mixture of saturated and aromatic hydrocarbons with a smaller fraction of heteroatom-containing compounds, i.e. the resins and asphaltenes. The molecular distribution typically ranges from 16 to 1000 amu.[4] The number of unique compounds is extensive and more than 240 000 molecular species have been resolved in a single sample.[5,6] Due to this complexity, a large portion of the petroleum chemical space is structurally unknown.

We have previously looked at the resins fraction (i.e. larger heteroatom-containing species) of North Sea oils.[7] Herein, we extend our work with a focus on aromatics. Within this fraction, the dominant species (in terms of abundance) are monoaromatic followed by a smaller amount of polycyclic aromatic hydrocarbons (PAHs).[8,9] The PAHs class is dominated by smaller (2 to 3 rings) PAHs, with larger species (e.g. chrysene, coronene) being present at trace levels. It also contains small amounts of heteroatomic-containing ring structures.[10–12] Due to their toxicity, PAHs have been extensively studied.[13] A large focus has been on the 16 priority pollutants PAHs defined by the U.S. Environmental Protection Agency.[14] However, this list is not representative of crude oils which contain a more structurally diverse PAH set.[15,16] Low molecular weight PAHs are susceptible to weathering, primarily by volatilization, whereas high molecular weight aromatics are more resilient.[17] Therefore, these are useful targets for oil-oil and oil-source correlation and spill identification and environmental monitoring.[18,19]

Comprehensive identification of the aromatics is challenging due to the large concentration variance. Traditionally, petroleum analysis is based on pre-fractionation using silica chromatography or solid-phase extraction (SPE) cartridges followed by GC-MS^n.[20–23] SPE is a low-efficiency separation technique, depending on chemical selectivity. This allows crude isolation of the aromatics fraction, but not separation of the compounds within it. Thus, an aromatic fraction obtained by SPE contains both the benzenes, naphthalenes and larger rings. Here, a typical crude oil will have a high abundance of monoaromatics, with diminishing concentrations with increasing ring size. The appropriate GC on-column concentration of the naphthalenes typically results in the larger ring systems being below the limit of detection (LOD). By increasing concentration to push trace aromatics above the LOD, both the column and detector will be saturated by the more abundant compounds. This leads to high background levels which affect quantitation and may obscure mass spectra complicating structural identification of unknowns.[24,25] Furthermore, the poor resolution of SPE often leads to an overlap between the saturated and aromatic hydrocarbons, which interferes with subsequent analysis. Thus, although SPE is efficient for routine applications, a large portion of the sample remains *invisible*. Comprehensive multi-dimensional chromatography (GC×GC) is often used as an alternative to simplify or remove the need for sample pre-fractionation.[26–28] However, it does not solve the issue with variable abundance and

column/detector overload. Thus, a complete qualitative and explorative oil analysis requires a more selective sample-prefractionation method.

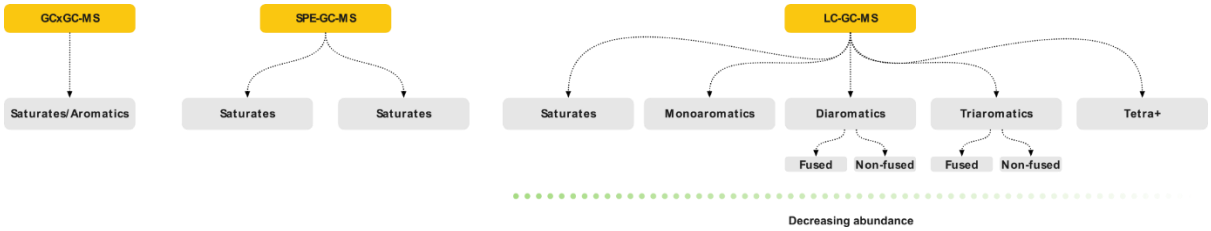

Figure 1. Schematic of the three sample preparation and analytical strategies discussed within the paper.

Herein, we present an HPLC-based method for the automated high-resolution fractionation of crude oil using commercially available columns. The method can resolve aromatics based on ring size and connectivity, i.e. fused and non-fused rings (*e.g. naphthalene versus biphenyl*). The fractions may be diluted or concentrated, depending on the target, for subsequent analysis and can thus be used to concentrate trace species. The method is easily modified to selectively collect only fractions of interest, and the aromatics may be collected either as one or several fractions. We demonstrate the method's applicability by fractionation of fourteen crude oils with subsequent GC-MS analysis. The method is compared to data obtained using SPE and GC×GC-MS. We demonstrate how it's especially suitable for the analysis and identification of trace aromatics by the successful tentative identification of several compounds not observed using comparable methods.

2 Materials and methods

2.1 Chemicals and reagents

Chloroform, dichloromethane, *n*-hexane, deuterated standards and model compounds (ethylbenzene, naphthalene, biphenyl, phenanthrene, 1-benzylnaphthalene and chrysene) were purchased from Sigma Aldrich and used as received.

2.2 Samples

Fourteen crude oils sampled from producing fields in the Danish region of the North Sea were obtained from Mærsk Oil (now Total E&P). The samples were received in metal containers (jerrycans) and transferred to glass bottles upon arrival. The samples were stored at room temperature protected from light.

2.3 Sample preparation

2.4 Solid-phase extraction

Crude oil (10 µL) was combined with 100 µL of a solution containing alkane internal standards (decane-d22, hexadecane-d34 and eicosane-d42, 400 µg/mL in n-hexane), 50 µL of PAH internal standards (naphthalene-d8, phenanthrene-d10, acenaphthene-d10, chrysene-d12 and perylene-d12, 30 µg/mL in n-hexane) and further diluted with *n*-hexane (840 µL). A solid-phase extraction column (Phenomenex EPH Strata, 200 µm, 70 Å, 500 mg / 3 mL) was cleaned and conditioned by CH_2Cl_2 (3x1 mL) followed by *n*-hexane (3x1 mL). 100 µL of oil solution was carefully applied to the column and was allowed to settle for 5 minutes.

Saturated hydrocarbons were eluted into one fraction with three portions of *n*-hexane (3x600 µL). Aromatic hydrocarbons were eluted using dichloromethane (1x1800 µL). The solvent level of each fraction was reduced to 500 µL under a gentle stream of nitrogen without applied heating to avoid losses of volatile components.

2.5 Liquid chromatography fractionation

Fractionation of crude oil was carried out on a Dionex UltiMate 3000 HPLC equipped with a DAD-3000 diode array, a RefractoMax RI-521 refractive index (RI) detector and an AFC-3000 fraction collector. The system was fitted with one six-port/two-way and one ten-port/two-way port to enable selective backflush of the primary column. A Thermo Scientific Hypersil Gold NH_2 (4.6 mm i.d., 3 µM, 150 mm) and a Hypersil Silica (4.6 mm i.d., 3 µM, 150 mm) were connected in series. The sample manager was kept at 20 °C and the column oven at 30 °C. The injection volume was 50 µL.

Samples were diluted at 1:2000 in *n*-hexane and stored at -20 °C for 24 hours to precipitate asphaltenes. The samples were centrifuged and an aliquot of the mother liquor was carefully transferred to an autosampler vial for analysis. Separation of saturates and aromatics was achieved via isocratic *n*-hexane elution during which 30 s wide fractions were collected. After elution of aromatics, the primary column was rinsed using a backflush gradient from *n*-hexane to 1:1 2-propanol:chloroform. The collected fractions were diluted (saturates, mono- and di-aromatics) or concentrated (tri-aromatics and larger) for analysis on GC-MS. For enrichment experiments, consecutive fractionations (typically 3 to 6) were performed with pooling of the eluents followed by solvent reduction under a gentle stream of N_2 at 30 °C.

2.6 Analytical methods

2.7 GC-MS

GC-MS data were recorded using an Agilent 5977B GC-MSD as follows; 250 °C inlet, 320 °C transfer line, splitless injection (1 µL), Agilent DB-5MS (30 m, 0.25 mm i.d., 0.25 µm). The oven temperature gradient was programmed as follows; 50 (1 min. hold-time) - 320 °C (8 min hold-time, 10 °C/min.), helium carrier gas at 1.5 mL/min. in constant flow mode.

GC×GC-MS data were recorded using an Agilent 7200B GC-QTOF equipped with a Zoex ZX-2 thermal modulator (Zoex Corporation, Houston, TX, USA) as follows; 250 °C inlet, 320 °C transfer line, splitless injection (1 µL), Agilent DB-5MS UI (1D, 30 m, 0.25 mm i.d., 0.25 µm df) and a Restek Rxi-17Sil MS (2D, 2 m, 0.18 mm i.d., 0.18 µm df) capillary columns connected using a SilTite µ-union. The oven was temperature programmed as follows; 50 (1 min hold-time) - 320 °C (3 °C/min.), helium carrier gas at 1 mL/min. in constant flow mode. The modulation period was set to 6 s with a 400 ms hot-jet duration.

2.8 Data processing

Data were screened using Masshunter Qualitative Navigator (Agilent, B.08.00). Peak detection and compound identification was performed using MassHunter Unknowns Analysis (Agilent, B.09.00) and the NIST Standard Reference Database (1A v17, Gaithersburg, MD, USA). Feature tables were exported as CSV files and imported into a Jupyter Notebook for

further processing using the Python scripting language.[29,30] Duplicates based on the CAS number were removed from the feature tables. All compounds containing silica and halogens were removed. The double-bond equivalent values were calculated for each compound and all features with a DBE of less than 4 were excluded. Finally, experimental and literature retention indices (RI) were compared with flagging of all compounds where the difference was larger than 50 units.

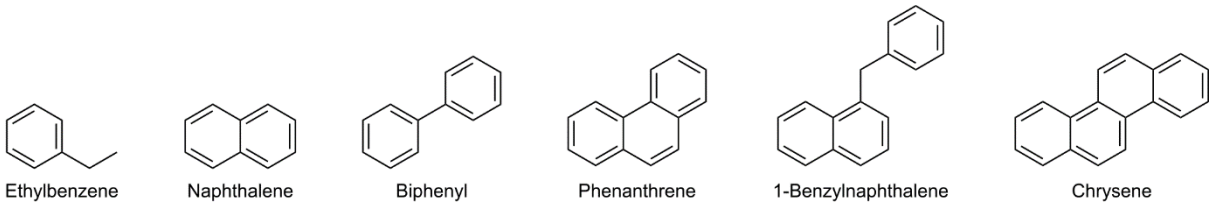

Figure 2. Molecular structure of the model compounds used for method development and validation.

3 Results and discussion

3.1 Experimental setup and method development

The objective of the method was to 1) separate saturates and aromatics and 2) intra-class separation of the aromatics with enrichment capabilities. A dual-column setup using normal phase analytical LC-columns provided the required selectivity and efficiency. The primary column (*Thermo Scientific Hypersil Gold* NH_2) acted as a retainer for polar components, whereas a secondary pure silica-based column (*Thermo Scientific Hypersil Silica*) was required for the separation of saturated and aromatic hydrocarbons. The separation was optimized using six model compounds commonly found in crude oil (Figure 2). An isocratic *n*-hexane elution yielded separation of saturated and mono-aromatic hydrocarbons, as well as separation of polycyclic aromatics based on ring size and connectivity (Figure 3).

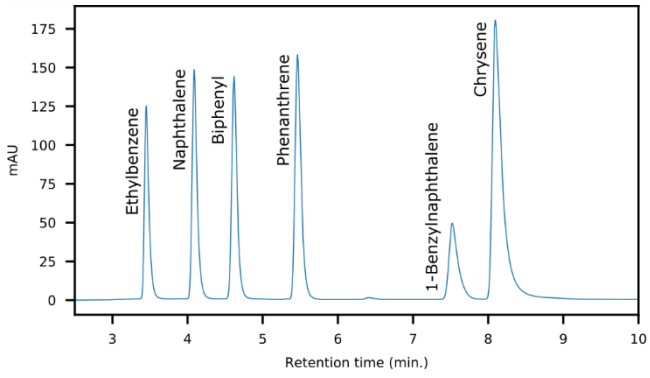

Figure 3. HPLC-Chromatogram showing the elution order and separation of model compounds.

The fraction collector was programmed to collect 30 second wide fractions based on the peak widths of the model compounds. At this fraction width, we observed only a minor overlap of fractions with co-elution of the most abundant components. A reduction of the fraction width can be set if a higher peak purity is required. The cost is a slight loss of recovery. After elution of the last aromatics as observed by UV/Vis, the flow path was selectively reversed

for the primary column. The column was then rinsed using a gradient from 100% *n*-hexane to 50:50 chloroform:2-propanol. This effectively removed the adsorbed resins on the amide column. The fraction collector is within the flow path during all stages of chromatography and the resins may therefore be isolated for further analysis[7]. The final step is a re-equilibration of the whole system by a return to isocratic *n*-hexane and flushing at an increased flow rate to remove the polar solvents from the flow path. Improper re-equilibration resulted in a severe loss of retention in subsequent fractionations due to the adsorption of 2-isopropanol on the silica phase.

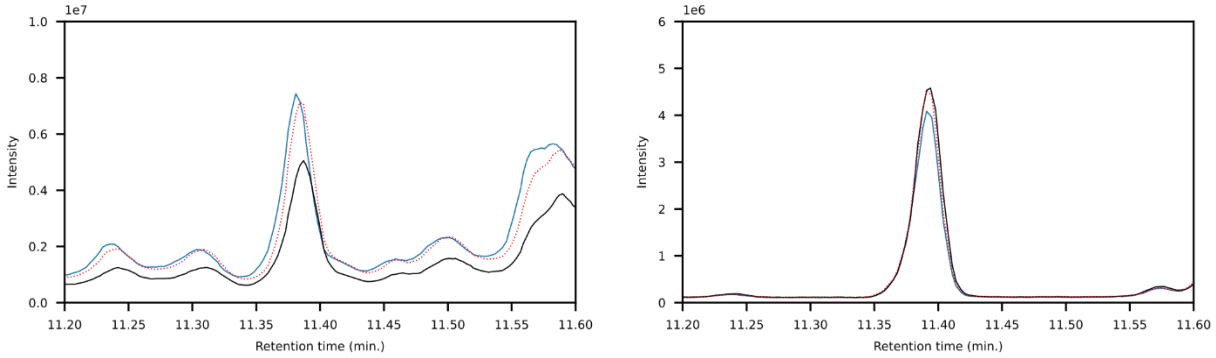

Figure 4. Chromatogram of the phenanthrene peak for triplicate SPE-extractions (left) compared to LC-fractionation (right).

Model compounds	**Area, STD GC-MS (%RSD, N = 6)**	**Area,STD LC-GC-MS (%RSD, N = 6)**	**Recovery (%)**
Naphthalene	3344437 (1.0%)	3566196 (5.7%)	106.6
Biphenyl	1525309 (1.6%)	1627554 (2.1%)	106.7
Phenanthrene	1145132 (3.1%)	1162530 (4.7%)	101.5
1-Benzylnaphthalene	599597 (4.4%)	592272 (3.0%)	98.8
Chrysene	220091(5.1%)	195914 (2.8%)	89.0
Crude oil compounds	**Area, SPE-GC-MS (%RSD, N = 3)***	**Area,oil LC-GC-MS (%RSD, N = 3)***	
1,2,4-Trimethylbenzene	29538362 (17.6%)	12638320 (4.6%)	
Naphthalene	20202722 (18.0%)	5153076 (8.1%)	
Phenanthrene	9239818 (17.4%)	3478436 (6.4%)	

Table 1. Recovery calculations based on integrated area as determined by GC-MS for pure model compounds and those isolated using HPLC-fractionation. * = The area difference is due to different dilution factors.

Recovery values were calculated using the model compounds by comparison of peak areas obtained on GC-MS from LC-fractions compared to direct analysis of the standards. Three analytes have recovery values slightly above 100%. This is likely due to a discrepancy between programmed and real injection volume on the HPLC auto-sampler. In contrast, the recovery is less than 90% for chrysene. For this compound, we observe peak broadening due to the high capacity factor. As the fraction collection width is static during the full run, the low recovery is attributed to the peak being wider than the collection width. To evaluate the reproducibility of complex samples, a single oil was fractionated three consecutive times. Each fraction was analyzed on GC-MS and the relative standard deviation was determined from peak areas. 1,2,4-Trimethylbenzene, naphthalene and phenanthrene gave 4.6,8.1 and 6.4% respectively. The results are similar to those obtained using a model mixture. This

shows that the method performs consistently in the presence of a highly complex oil matrix. Furthermore, the method successfully removes interferences and yields a high signal-to-noise ratio for the target analytes in each fraction (Figure 4).

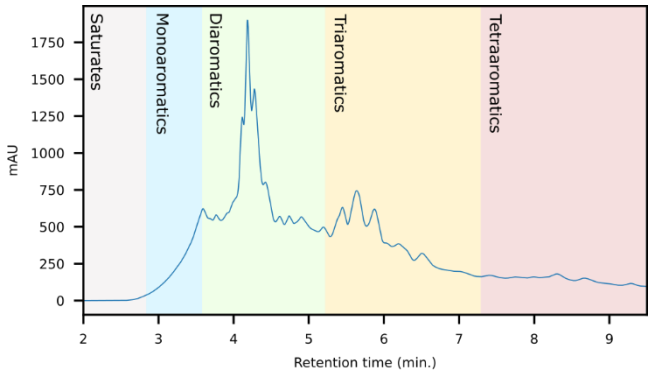

Figure 5. Chromatogram (UV response at 272 nm) of a representative crude oil with approximate retention regions of the aromatic classes.

3.2 Applicability in crude oil analyses

The applicability of the method was demonstrated by fractionation and analysis of 14 crude oils. The oil samples were sourced from producing fields in the Danish region of the North Sea. Crude oils from this region typically have an aromatics content of 25 – 30%, of which the majority are BTEX-type monoaromatics (*benzene*, *toluene*, *xylene*) with a continuous decrease in abundance with increasing ring size[7]. The primary fraction is the saturated hydrocarbons followed by the resins (up to 5%) with only traces of asphaltenes. This is evident from the fractionation, where a typical dilution factor of 50/20 had to be applied to the saturated and monoaromatic fractions respectively. The fractions containing larger aromatics were analyzed either undiluted or concentrated by solvent reduction.

The first fraction contains the paraffins and naphthenes and is poorly retained on the primary LC-column (Figure 4). The secondary second silica column is required to separate them from the monoaromatics, which elute as the second fraction (Figure 5). The third and fourth fractions contain diaromatic species, with the latter non-fused ring systems (e.g. naphthalene versus biphenyl). Fractions 5 and 6 contain the triaromatics (e.g. phenanthrene versus 1-phenylnaphthalene). Here, the abundance starts to diminish and the sixth fraction had to be concentrated for subsequent GC-MS analysis. Fractions 7 and above contain larger ring systems, e.g. chrysene, perylene. These fractions are less well-defined, likely because compounds eluting within this retention range are fewer in number and present in trace amounts. We also observed a slight loss of resolution, with minor overlap and cross-contamination. This is a result of two things; 1) diffusion and peak broadening during the liquid chromatography 2) collection of low abundance (undiluted/concentrated) fraction after a high abundance (diluted) fraction. If higher purity peaks are required the fraction collection width can be reduced. Attempts to concentrate fractions 9 and later were not successful and

the gas chromatograms were dominated by background contaminants likely originating from the solvents, HPLC-tubing and glassware (e.g. siloxanes, surfactants).

Fraction	Main composition	Comment
1	Saturated hydrocarbons	High concentration; diluted for GC-MS.
2	Mono-aromatics	High concentration; diluted for GC-MS.
3	Di-aromatics	Medium concentration; diluted for GC-MS.
4	Non-fused di-aromatics	Medium concentration; diluted for GC-MS.
5	Tri-aromatics	Low concentration; undiluted for GC-MS.
6	Non-fused tri-aromatics	Trace concentrations; concentrated for GC-MS.
7 + 8	Misc. tetra-aromatics	Trace concentrations; concentrated for GC-MS.

Table 2. Summary description of the major constituents of each fraction.¨

3.3 Performance comparison with SPE-GC-MS and GC×GC-MS

Solid-phase extraction of crude oil into its saturated and aromatic fraction is a well-established sample preparation method. The physical properties of SPE adsorbents (large particle size, low mass loadings) result in limited separation power.[31,32] Thus, the technique mainly applicable for the crude separation of different compound classes. It does not provide sufficient resolution to separate closely related compounds within subfractions. To compare our method to SPE we fractionated each oil using a Phenomenex Strata EPH (200 µm, 70 Å, 500 mg / 3 mL). The cartridge contains a proprietary phase specifically developed to separate hydrocarbon fractions.[33]

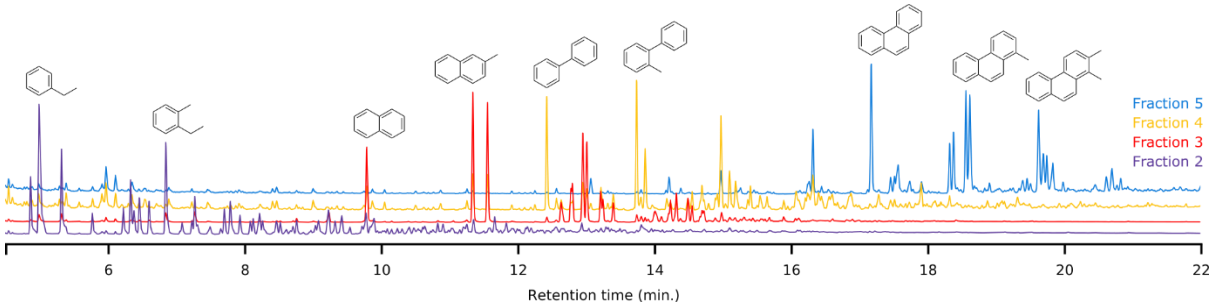

Figure 5. GC-MS chromatograms of fractions two to five with identification of representative compounds.

In terms of spectral quality, an approximately 10-fold reduction in background noise is observed in the LC fractions as compared to SPE. A comparison of the extracted mass spectra for the peak corresponding to 1,3-dimethylpyrene is presented in Figure 6. Selected ion monitoring can be used to reduce background interferences for target species but results in loss of spectral detail for qualitative analysis. Furthermore, when using low-resolution instruments, i.e. single quadrupole MS, there is a large risk of overlap in complex samples.[34,35] The reduction in background noise improved library matching, especially for analytes present at trace levels.

Filter	GCxGC Tot.	GCxGC Unique	SPE-GC Tot.	SPE-GC Unique	LC-GC Tot.	LC-GC Unique
No formula or R.I filter	181	98	601	191	957	535
No formula filter, R.I ±100 units	82	15	300	68	426	187
N,S,O≥1, no R.I filter	80	79	311	151	517	357
N,S,O≥1, R.I ±100 units	6	6	91	44	124	81

Table 5. Comparison of the number of tentatively identified species in SPE-GC versus LC-GC. The calculations are based on merged data of all samples with the removal of duplicates. Unique indicates compounds not identified in the compared methods.

To evaluate identification performance, peak picking and library matching were carried out using MassHunter Unknowns Analysis and the NIST mass spectral library. The match factor limit was set to 700. The number of compounds was compared both on a sample-to-sample basis and by merging all features from all samples (with duplicates removal based on CAS number). A comparison of the merged compound tables of all samples shows that using the LC-GC method we can identify 957 compared to 601 compounds using SPE. This is an increase of 37.2%. To increase the match confidence, we applied a retention index (RI) filter, only retaining compounds with a match within 100 units of the library value. By doing so, we identified 426 compared to 300 (42% increase). This excludes all compounds of which a library RI is not available (approximately 1% of our feature set). However, a large portion of the compounds only have computationally approximated retention indices and not experimentally determined values. Thus, all filtering and data analyses should be carried out with care and manual intervention. The SPE fraction contains approximately 190 unique compounds with 404 compounds overlapping both analyses. Manual inspection reveals this list contains several petroleum-type compounds and not predominantly background noise or contamination (e.g. plasticizers, column contamination). One plausible source is errors occurring during the automatic processing routines. Small differences in mass spectra (e.g. due to abundance or background level) can lead to closely related library matches being given similar (but different) priority (e.g. isomeric species). Figure 7 (left) shows the DBE distributions of assigned compounds uniquely observed from SPE-GC and LC-GC methods. Noticeably, the DBE distributions are significantly different between the two methods. The distribution of unique compounds from SPE-GC is centered around low DBE 4 and 5 (e.g. monoaromatic), whereas unique compounds from LC-GC are distributed more evenly at higher DBE values. This is expected as the LC-GC method isolates and enriches high aromaticity fractions. These findings showcase the ability of LC-GC as a high-resolution fractionation method for crude oil.

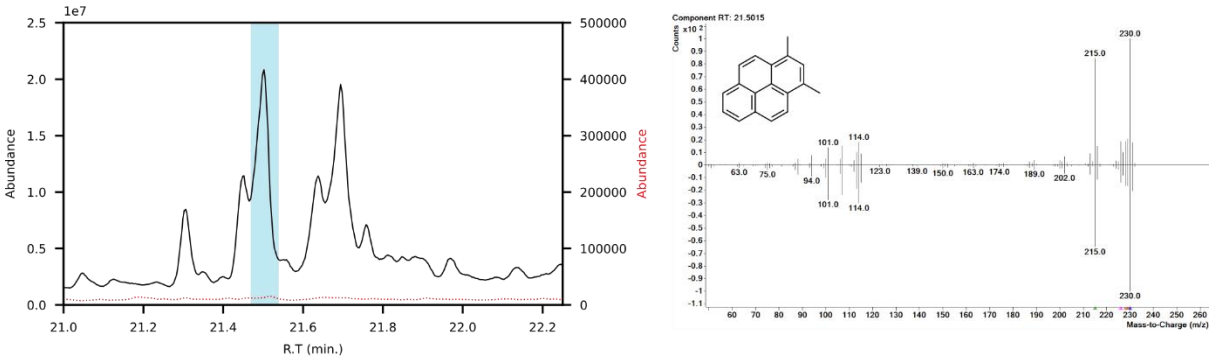

Figure 6. *Left;* Chromatogram overlay showing the peak of 1,3-dimethylpyrene found in LC-fraction 7 (*black, pre-concentrated*) and the aromatics fraction obtained via SPE (dotted red). *Right;* Mass spectral mirror plot of the LC-fraction peak showcasing its purity.

For comparison to comprehensive multi-dimensional chromatography, a single sample was analyzed by our in-house routine GC×GC-MS method (i.e. solvent dilution, filtration and analysis) (Figure 8). The objective of the GC×GC method is not to maximize feature ID but enable multi-class analysis/fingerprinting with minimal to no sample preparation. Furthermore, the SPE/LC-GC and GC×GC analyses were carried out on different instruments which makes direct comparison challenging. For GC×GC an Agilent 7200B QTOF high-resolution mass spectrometer was used. For SPE/LC-GC, an Agilent 5977B single quadrupole equipped with a High-Efficiency Source (HES) was used. The HES has both higher sensitivity and dynamic range. Secondly, for GC×GC, the dilution factor was adjusted so that the analytes with the highest abundance were at detector saturation. Here, we see that although GC×GC is not restricted in terms of peak capacity, it does fall short in terms of dynamic range. After blob detection, library matching and filtering we obtain 63 tentative hits in a single sample. With corresponding processing settings, we identified 143 compounds by multi-fraction LC-GC-MS analysis of the same sample. This is an increase of 127%. In terms of manual intervention, ease of use and time of analysis, GC×GC is preferred compared to LC-GC-MS. However, the amount of data generated using the latter is more comprehensive in our case.

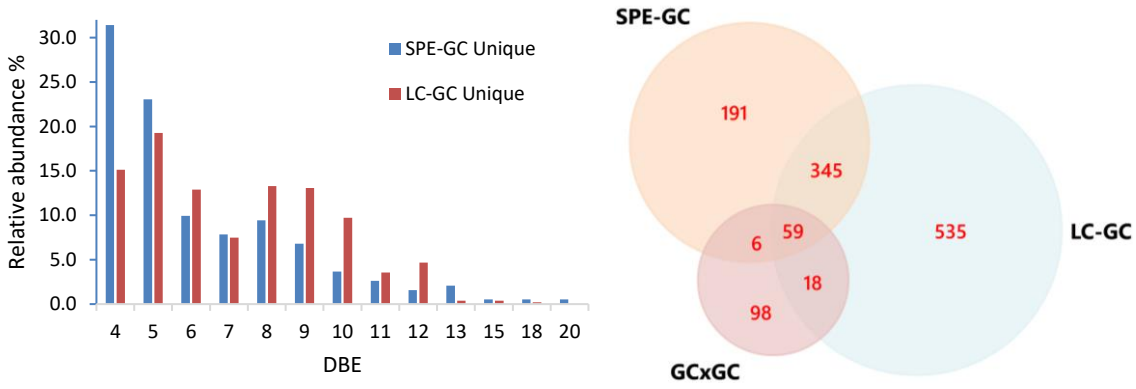

Figure 7. *Left*; Double bond equivalent (DBE) distribution plot for assigned unique compounds obtained from SPE-GC and LC-GC methods. *Right*; Venn diagram showing the total amount of assigned compositions obtained from SPE-GC-, LC-GC- and GCxGC-MS.

A Venn diagram was constructed to compare three methods (Figure 7, right). The number within each colored circle represents the number of assigned unique compounds for each method, whereas numbers in overlapped zones represent the number of compounds that have been co-assigned from corresponding methods. The amount of compositions obtained from LCxGC (957) significantly surpasses GCxGC (181) and SPE-GC (601). We obtained approximately 50% unique compounds with GCxGC and LC-GC and 32% with SPE-GC. The Venn diagram also shows that co-assigned compounds of those three methods cover a narrow range of overall chemical composition (59 co-assigned compounds) of crude oil. Again, it is worth noting that there are differences in terms of dilution factor and instrumental parameters for those methods. Therefore, the comparison is biased but still relevant to evaluate the LC-GC method for trace components analysis.

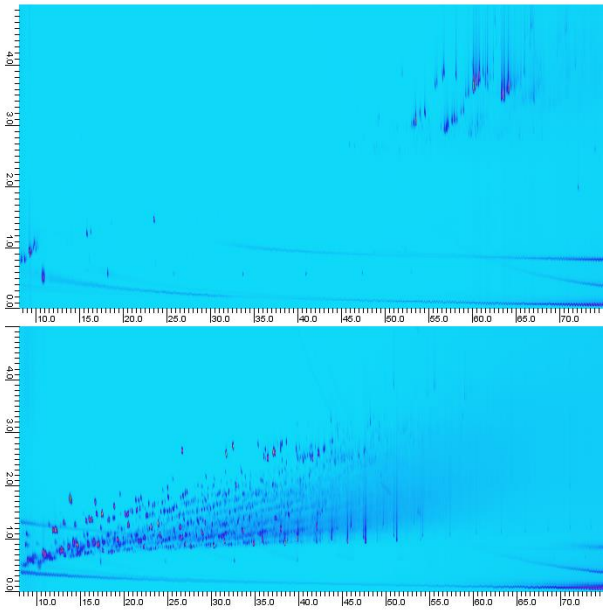

Figure 8. Comparison of whole oil GC×GC- (bottom) and fraction 5 obtained from LC-GC-MS (top). The LC-GC fraction has a high abundance cluster of alkyl phenanthrene isomers at or below the LOD by whole oil GC×GC.

4 Concluding remarks

We have developed a method for high-resolution fractionation of complex crude oil matrices. By using sub-micron LC columns we obtained high efficiency and resolution which allowed intra-class compound separation. This is in contrast with traditional methods, e.g. SPE, which yields a single aromatics fraction. The method is especially advantageous for the isolation of trace species. Multiple compounds not observed by SPE-GC-MS were pre-concentrated

yielding high abundance and spectral quality. The increase in the number of tentatively identified peaks is thus a result of both reduced co-elution and an increase in analyte signal-to-noise ratio.

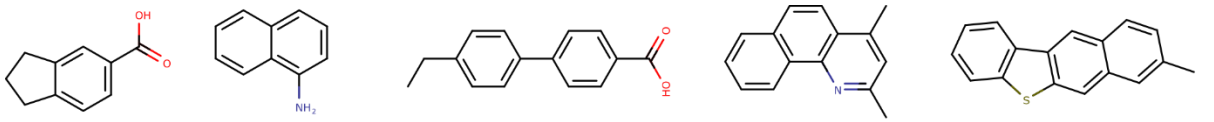

Figure 9. The molecular structures of five representative compounds identified in HPLC-fractions 2 to 6.

By characterization of 14 crude oils, we extended the identification to a large number of hydrocarbon and N,S,O-containing aromatics. Of the 517 uniquely identified compounds, 69% (357) contain either N,S,O (or a combination of) atoms (Table 3). The structures of five representative compounds are presented in Figure 9. Aromatic nitrogen and sulfur compounds are detrimental in petroleum processing. Furthermore, they potentially have biological activity and may pose an environmental and toxicological hazard.[11,36–38] Therefore, their characterization is an important pursuit. They are routinely analyzed by direct infusion mass spectrometry that provides the molecular formula but not connectivity.[39–42] Thus, isolation and GC-MS analysis with library matching provide valuable information on their presence in oil samples.

The relatively long fractionation time (60 minutes) and the number of fractions generated lead to a full sample analysis time of 6 hours (when characterizing the first 7 fractions using GC-MS). Several steps require manual intervention, i.e. dilution and pre-concentration of fractions and moving the samples from the LC to the GC. It would therefore be beneficial to implement more automation, e.g. by using liquid handling robotics (ultimately with direct hyphenation to the GC). We observed minor co-elution during the analysis of latter fractions. Combining the LC-fractionation with subsequent GC×GC analysis would increase the power of the method further. However, it would require an intense data processing workflow with high demands in computational power. Something that is already challenging in comprehensive GC×GC studies.[43–45]

5 Author contributions

J.S. conceptualized the study. A.K, A.E.J and J.S developed the sample preparation workflow and analytical methodology. J.S carried out the LC-based fractionation and analysis. A.E.J carried out the solid-phase extractions and analysis. A.K carried out the GC×GC-analysis. J.S and A.K. developed the data processing workflow and analyzed the data. A.K. and J.S wrote the manuscript. All authors performed scientific and grammatical revisions to the final draft.

6 Acknowledgments

The authors are grateful for the financial support received from the Danish Hydrocarbon Research & Technology Centre (DHRTC) under the Reservoir Fluids Characterization

project. The authors wish to thank Karen L. Feilberg for discussions. The authors are also grateful for the donation of the samples from Total E&P Denmark. Furthermore, the authors kindly acknowledge the Danish Underground Consortium (Total, E&P Denmark, Noreco & Nordsøfonden) for granting permission to publish this work.

7 Conflict of interest statement

The authors declare that there is no conflict of interest.

References

(1) Net Zero by 2050 A Roadmap for the Global Energy Sector. International Energy Agency 2021.

(2) Aftalion, F. *A History of the International Chemical Industry*, 2nd ed.; Chemical Heritage Press: Philadelphia, PA, 2001.

(3) Yadav, V. G.; Yadav, G. D.; Patankar, S. C. The Production of Fuels and Chemicals in the New World: Critical Analysis of the Choice between Crude Oil and Biomass Vis-à-Vis Sustainability and the Environment. *Clean Techn Environ Policy* **2020**, *22* (9), 1757–1774. https://doi.org/10.1007/s10098-020-01945-5.

(4) Marshall, A. G.; Rodgers, R. P. Petroleomics: Chemistry of the Underworld. *Proceedings of the National Academy of Sciences* **2008**, *105* (47), 18090–18095. https://doi.org/10.1073/pnas.0805069105.

(5) Palacio Lozano, D. C.; Gavard, R.; Arenas-Diaz, J. P.; Thomas, M. J.; Stranz, D. D.; Mejía-Ospino, E.; Guzman, A.; Spencer, S. E. F.; Rossell, D.; Barrow, M. P. Pushing the Analytical Limits: New Insights into Complex Mixtures Using Mass Spectra Segments of Constant Ultrahigh Resolving Power. *Chem. Sci.* **2019**, *10* (29), 6966–6978. https://doi.org/10.1039/C9SC02903F.

(6) Krajewski, L. C.; Rodgers, R. P.; Marshall, A. G. 126 264 Assigned Chemical Formulas from an Atmospheric Pressure Photoionization 9.4 T Fourier Transform Positive Ion Cyclotron Resonance Mass Spectrum. *Anal. Chem.* **2017**, *89* (21), 11318–11324. https://doi.org/10.1021/acs.analchem.7b02004.

(7) Sundberg, J.; Feilberg, K. L. Characterization of Heteroatom Distributions in the Polar Fraction of North Sea Oils Using High-Resolution Mass Spectrometry. *Journal of Petroleum Science and Engineering* **2020**, *184*, 106563. https://doi.org/10.1016/j.petrol.2019.106563.

(8) Wei, O.; Xu, X.; Zhang, Y.; Yang, B.; Ye, Q.; Yang, Z. Multidimensional Gas Chromatography–Mass Spectrometry Method for Fingerprinting Polycyclic Aromatic Hydrocarbons and Their Alkyl-Homologs in Crude Oil. *Analytical Letters* **2018**, *51* (4), 483–495. https://doi.org/10.1080/00032719.2017.1341905.

(9) Requejo, A. G.; Sassen, R.; McDonald, T.; Denoux, G.; Kennicutt, M. C.; Brooks, J. M. Polynuclear Aromatic Hydrocarbons (PAH) as Indicators of the Source and Maturity of Marine Crude Oils. *Organic Geochemistry* **1996**, *24* (10–11), 1017–1033. https://doi.org/10.1016/S0146-6380(96)00079-4.

(10) Mössner, S. G.; Wise, S. A. Determination of Polycyclic Aromatic Sulfur Heterocycles in Fossil Fuel-Related Samples. *Anal. Chem.* **1999**, *71* (1), 58–69. https://doi.org/10.1021/ac980664f.

(11) Zhang, G.; Yang, C.; Serhan, M.; Koivu, G.; Yang, Z.; Hollebone, B.; Lambert, P.; Brown, C. E. Characterization of Nitrogen-Containing Polycyclic Aromatic Heterocycles in Crude Oils and Refined Petroleum Products. In *Advances in Marine*

Biology; Elsevier, 2018; Vol. 81, pp 59–96. https://doi.org/10.1016/bs.amb.2018.09.006.

(12) Carvalho Dias, L.; Bomfim Bahia, P. V.; Nery do Amaral, D.; Machado, M. E. Nitrogen Compounds as Molecular Markers: An Overview of Analytical Methodologies for Its Determination in Crude Oils and Source Rock Extracts. *Microchemical Journal* **2020**, *157*, 105039. https://doi.org/10.1016/j.microc.2020.105039.

(13) Lawal, A. T. Polycyclic Aromatic Hydrocarbons. A Review. *Cogent Environmental Science* **2017**, *3* (1), 1339841. https://doi.org/10.1080/23311843.2017.1339841.

(14) Keith, L. H. The Source of U.S. EPA's Sixteen PAH Priority Pollutants. *Polycyclic Aromatic Compounds* **2015**, *35* (2–4), 147–160. https://doi.org/10.1080/10406638.2014.892886.

(15) Stout, S. A.; Emsbo-Mattingly, S. D.; Douglas, G. S.; Uhler, A. D.; McCarthy, K. J. Beyond 16 Priority Pollutant PAHs: A Review of PACs Used in Environmental Forensic Chemistry. *Polycyclic Aromatic Compounds* **2015**, *35* (2–4), 285–315. https://doi.org/10.1080/10406638.2014.891144.

(16) Andersson, J. T.; Achten, C. Time to Say Goodbye to the 16 EPA PAHs? Toward an Up-to-Date Use of PACs for Environmental Purposes. *Polycyclic Aromatic Compounds* **2015**, *35* (2–4), 330–354. https://doi.org/10.1080/10406638.2014.991042.

(17) John, G. F.; Han, Y.; Clement, T. P. Weathering Patterns of Polycyclic Aromatic Hydrocarbons Contained in Submerged Deepwater Horizon Oil Spill Residues When Re-Exposed to Sunlight. *Science of The Total Environment* **2016**, *573*, 189–202. https://doi.org/10.1016/j.scitotenv.2016.08.059.

(18) *Petrogenic Polycyclic Aromatic Hydrocarbons in the Aquatic Environment: Analysis, Synthesis, Toxicity and Environmental Impact*; M. Pampanin, D., O. Sydnes, M., Eds.; BENTHAM SCIENCE PUBLISHERS, 2017. https://doi.org/10.2174/9781681084275117010l.

(19) Poulsen, K. G.; Kristensen, M.; Tomasi, G.; Dela Cruz, M.; Christensen, J. H. The Pixel-Based Chemometric Approach for Oil Spill Identification and Hydrocarbon Source Differentiation. In *Oil Spill Environmental Forensics Case Studies*; Elsevier, 2018; pp 443–463. https://doi.org/10.1016/B978-0-12-804434-6.00021-5.

(20) Gilgenast, E.; Boczkaj, G.; Przyjazny, A.; Kamiński, M. Sample Preparation Procedure for the Determination of Polycyclic Aromatic Hydrocarbons in Petroleum Vacuum Residue and Bitumen. *Anal Bioanal Chem* **2011**, *401* (3), 1059–1069. https://doi.org/10.1007/s00216-011-5134-9.

(21) Pillai, I.; Ritchie, L.; Heywood, R.; Wilson, G.; Pahlavanpour, B.; Setford, S.; Saini, S. Development of an Improved Analytical Method for the Determination of Carcinogenic Polycyclic Aromatic Hydrocarbons in Transformer Oil. *Journal of Chromatography A* **2005**, *1064* (2), 205–212. https://doi.org/10.1016/j.chroma.2004.12.066.

(22) Wang, Z.; Fingas, M.; Li, K. Fractionation of a Light Crude Oil and Identification and Quantitation of Aliphatic, Aromatic, and Biomarker Compounds by GC-FID and GC-MS, Part I. *Journal of Chromatographic Science* **1994**, *32* (9), 361–366. https://doi.org/10.1093/chromsci/32.9.361.

(23) Alzaga, R.; Montuori, P.; Ortiz, L.; Bayona, J. M.; Albaigés, J. Fast Solid-Phase Extraction–Gas Chromatography–Mass Spectrometry Procedure for Oil Fingerprinting. *Journal of Chromatography A* **2004**, *1025* (1), 133–138. https://doi.org/10.1016/j.chroma.2003.10.080.

(24) Zhao, Y.; Hong, B.; Fan, Y.; Wen, M.; Han, X. Accurate Analysis of Polycyclic Aromatic Hydrocarbons (PAHs) and Alkylated PAHs Homologs in Crude Oil for

Improving the Gas Chromatography/Mass Spectrometry Performance. *Ecotoxicology and Environmental Safety* **2014**, *100*, 242–250. https://doi.org/10.1016/j.ecoenv.2013.10.018.
(25) Wilton, N. M.; Wise, S. A.; Robbat, A. Errors in Alkylated Polycyclic Aromatic Hydrocarbon and Sulfur Heterocycle Concentrations Caused by Currently Employed Standardized Methods. *Analytica Chimica Acta* **2017**, *977*, 20–27. https://doi.org/10.1016/j.aca.2017.04.017.
(26) Stilo, F.; Bicchi, C.; Jimenez-Carvelo, A. M.; Cuadros-Rodriguez, L.; Reichenbach, S. E.; Cordero, C. Chromatographic Fingerprinting by Comprehensive Two-Dimensional Chromatography: Fundamentals and Tools. *TrAC Trends in Analytical Chemistry* **2021**, *134*, 116133. https://doi.org/10.1016/j.trac.2020.116133.
(27) Jennerwein, M. K.; Eschner, M.; Gröger, T.; Wilharm, T.; Zimmermann, R. 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* **2014**, *28* (9), 5670–5681. https://doi.org/10.1021/ef501247h.
(28) Nizio, K. D.; McGinitie, T. M.; Harynuk, J. J. Comprehensive Multidimensional Separations for the Analysis of Petroleum. *Journal of Chromatography A* **2012**, *1255*, 12–23. https://doi.org/10.1016/j.chroma.2012.01.078.
(29) Kluyver, T.; Ragan-Kelley, B.; Pérez, F.; Granger, B.; Bussonnier, M.; Frederic, J.; Kelley, K.; Hamrick, J.; Grout, J.; Corlay, S.; Ivanov, P.; Avila, D.; Abdalla, S.; Willing, C. Jupyter Notebooks – a Publishing Format for Reproducible Computational Workflows. In *Positioning and Power in Academic Publishing: Players, Agents and Agendas*; Loizides, F., Schmidt, B., Eds.; IOS Press, 2016; pp 87–90.
(30) *Python Language Reference*; Python Software Foundation.
(31) Berrueta, L. A.; Gallo, B.; Vicente, F. A Review of Solid Phase Extraction: Basic Principles and New Developments. *Chromatographia* **1995**, *40* (7–8), 474–483. https://doi.org/10.1007/BF02269916.
(32) Buszewski, B.; Szultka, M. Past, Present, and Future of Solid Phase Extraction: A Review. *Critical Reviews in Analytical Chemistry* **2012**, *42* (3), 198–213. https://doi.org/10.1080/07373937.2011.645413.
(33) Countryman, S.; Kelly, K.; Garriques, M. An Improved Method for Diesel Fractionation Following the Revised Massachusetts EPH Guidelines. Phenomenex Inc., Torrance, CA, USA.
(34) Rosenthal, David. Theoretical Limitations of Gas Chromatographic/Mass Spectrometric Identification of Multicomponent Mixtures. *Anal. Chem.* **1982**, *54* (1), 63–66. https://doi.org/10.1021/ac00238a020.
(35) Davis, J. M.; Giddings, J. Calvin. Statistical Theory of Component Overlap in Multicomponent Chromatograms. *Anal. Chem.* **1983**, *55* (3), 418–424. https://doi.org/10.1021/ac00254a003.
(36) Anyanwu, I. N.; Semple, K. T. Fate and Behaviour of Nitrogen-Containing Polycyclic Aromatic Hydrocarbons in Soil. *Environmental Technology & Innovation* **2015**, *3*, 108–120. https://doi.org/10.1016/j.eti.2015.02.006.
(37) Vetere, A.; Pröfrock, D.; Schrader, W. Qualitative and Quantitative Evaluation of Sulfur-Containing Compound Types in Heavy Crude Oil and Its Fractions. *Energy Fuels* **2021**, *35* (10), 8723–8732. https://doi.org/10.1021/acs.energyfuels.1c00491.
(38) López García, C.; Becchi, M.; Grenier-Loustalot, M. F.; Païsse, O.; Szymanski, R. Analysis of Aromatic Sulfur Compounds in Gas Oils Using GC with Sulfur Chemiluminescence Detection and High-Resolution MS. *Anal. Chem.* **2002**, *74* (15), 3849–3857. https://doi.org/10.1021/ac011190e.

(39) Purcell, J. M.; Juyal, P.; Kim, D.-G.; Rodgers, R. P.; Hendrickson, C. L.; Marshall, A. G. Sulfur Speciation in Petroleum: Atmospheric Pressure Photoionization or Chemical Derivatization and Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Energy Fuels* **2007**, *21* (5), 2869–2874. https://doi.org/10.1021/ef700210q.

(40) Corilo, Y. E.; Rowland, S. M.; Rodgers, R. P. Calculation of the Total Sulfur Content in Crude Oils by Positive-Ion Atmospheric Pressure Photoionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Energy Fuels* **2016**, *30* (5), 3962–3966. https://doi.org/10.1021/acs.energyfuels.6b00497.

(41) Guan, S.; Marshall, A. G.; Scheppele, S. E. Resolution and Chemical Formula Identification of Aromatic Hydrocarbons and Aromatic Compounds Containing Sulfur, Nitrogen, or Oxygen in Petroleum Distillates and Refinery Streams. *Anal. Chem.* **1996**, *68* (1), 46–71. https://doi.org/10.1021/ac9507855.

(42) Purcell, J. M.; Rodgers, R. P.; Hendrickson, C. L.; Marshall, A. G. Speciation of Nitrogen Containing Aromatics by Atmospheric Pressure Photoionization or Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *J Am Soc Mass Spectrom* **2007**, *18* (7), 1265–1273. https://doi.org/10.1016/j.jasms.2007.03.030.

(43) Wilde, M. J.; Zhao, B.; Cordell, R. L.; Ibrahim, W.; Singapuri, A.; Greening, N. J.; Brightling, C. E.; Siddiqui, S.; Monks, P. S.; Free, R. C. Automating and Extending Comprehensive Two-Dimensional Gas Chromatography Data Processing by Interfacing Open-Source and Commercial Software. *Anal. Chem.* **2020**, *92* (20), 13953–13960. https://doi.org/10.1021/acs.analchem.0c02844.

(44) Stefanuto, P.-H.; Smolinska, A.; Focant, J.-F. Advanced Chemometric and Data Handling Tools for GC×GC-TOF-MS. *TrAC Trends in Analytical Chemistry* **2021**, *139*, 116251. https://doi.org/10.1016/j.trac.2021.116251.

(45) Reichenbach, S. E.; Tao, Q.; Cordero, C.; Bicchi, C. A Data-Challenge Case Study of Analyte Detection and Identification with Comprehensive Two-Dimensional Gas Chromatography with Mass Spectrometry (GC×GC-MS). *Separations* **2019**, *6* (3), 38. https://doi.org/10.3390/separations6030038.