Peer Chemistry Journals
Analytical, Inorganic, Organic, Physical, Materials Science

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

Khoa Huynh [1], **Annette E. Jensen** [1], **Jonas Sundberg** [Corresp. 1]

[1] Danish Hydrocarbon Research and Technology Centre, Technical University of Denmark, Kgs. Lyngby, Denmark

Corresponding Author: Jonas Sundberg
Email address: jonsun@dtu.dk

Characterization of crude oil remains a challenge for analytical chemists. With the development of multi-dimensional chromatography and high-resolution mass spectrometry, an impressive number of compounds can be identified in a single sample. However, the large diversity in structure and abundance makes it difficult to obtain full compound coverage. Sample preparation methods such as solid-phase extraction and SARA-type separations are used to fractionate oil into compound classes. 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. The method is based on semi-preparative liquid chromatography. This yields high selectivity and efficiency with separation based on aromaticity, ring size and connectivity. By the separation of the more abundant aromatics, i.e. monoaromatics and naphthalenes, trace species were isolated and enriched. This enabled 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 identify an extended set of heteroatom-containing aromatics (e.g. 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 Extended characterization of petroleum aromatics
2 using off-line LC-GC-MS

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

6 [1] Danish Hydrocarbon Research and Technology Centre, Technical University of
7 Denmark
8 Elektrovej 375, 2800 Kgs. Lyngby, Denmark.

10 Corresponding Author:
11 Jonas Sundberg[1]
12 Danish Hydrocarbon Research and Technology Centre, Technical University of
13 Denmark
14 Elektrovej 375, 2800 Kgs. Lyngby, Denmark. Email address: jonsun@dtu.dk

16 Abstract

17 Characterization of crude oil remains a challenge for analytical chemists. With the
18 development of multi-dimensional chromatography and high-resolution mass
19 spectrometry, an impressive number of compounds can be identified in a single sample.
20 However, the large diversity in structure and abundance makes it difficult to obtain full
21 compound coverage. Sample preparation methods such as solid-phase extraction and
22 SARA-type separations are used to fractionate oil into compound classes. However, the
23 molecular diversity within each fraction is still highly complex. Thus, in the routine
24 analysis, only a small part of the chemical space is typically characterized. Obtaining a
25 more detailed composition of crude oil is important for production, processing and
26 environmental aspects. We have developed a high-resolution fractionation method for
27 isolation and preconcentration of trace aromatics, including oxygenated and nitrogen-
28 containing species. The method is based on semi-preparative liquid chromatography.
29 This yields high selectivity and efficiency with separation based on aromaticity, ring size
30 and connectivity. By the separation of the more abundant aromatics, i.e. monoaromatics
31 and naphthalenes, trace species were isolated and enriched. This enabled the
32 identification of features not detectable by routine methods. We demonstrate the
33 applicability by fractionation and subsequent GC-MS analysis of 14 crude oils sourced
34 from the North Sea. The number of tentatively identified compounds increased by
35 approximately 60 to 150% compared to solid-phase extraction and GC×GC-MS.
36 Furthermore, the method was used to successfully identify an extended set of
37 heteroatom-containing aromatics (e.g. amines, ketones). The method is not intended to
38 replace traditional sample preparation techniques or multi-dimensional chromatography

39 but acts as a complementary tool. An in-depth comparison to routine characterization
40 techniques is presented concerning advantages and disadvantages.
41

Introduction

43 The use of petroleum as a feedstock for energy production is declining. However,
44 certain critical functions cannot safely be replaced by renewable energy ("Net Zero by
45 2050 A Roadmap for the Global Energy Sector," 2021). Secondly. petroleum is a
46 fundamental feedstock for the production of a large number of chemical starting
47 materials (Aftalion, 2001; Yadav, Yadav & Patankar, 2020). Therefore, reducing the
48 environmental impact of oil production is an important goal. This requires a better
49 understanding of petroleum on the molecular level. Crude oil is a complex mixture of
50 saturated and aromatic hydrocarbons with a smaller fraction of heteroatom-containing
51 compounds, i.e. the resins and asphaltenes. The molecular distribution typically ranges
52 from 16 to 1000 amu (Marshall & Rodgers, 2008). The number of unique compounds is
53 extensive and more than 240 000 molecular species have been resolved in a single
54 sample (Krajewski, Rodgers & Marshall, 2017; Palacio Lozano et al., 2019). Due to this
55 complexity, a large portion of the petroleum chemical space is structurally unknown.
56
57 We have previously looked at the resins fraction (i.e. polar heteroatom-containing
58 species) of North Sea oils (Sundberg & Feilberg, 2020). Herein, we extend our work
59 with a focus on aromatics. Within this fraction, the dominant species (in terms of
60 abundance) are monoaromatic followed by a smaller amount of polycyclic aromatic
61 hydrocarbons (PAHs) (Requejo et al., 1996; Wei et al., 2018). The PAHs class is
62 dominated by smaller (2 to 3 rings) PAHs, with larger species (e.g. chrysene, coronene)
63 being present at trace levels. It also contains small amounts of heteroatomic-containing
64 ring structures (Mössner & Wise, 1999; Zhang et al., 2018; Carvalho Dias et al., 2020).
65 Due to their toxicity, PAHs have been extensively studied (Lawal, 2017). A large focus
66 has been on the 16 priority pollutants PAHs defined by the U.S. Environmental
67 Protection Agency (Keith, 2015). However, this list is not representative of crude oils
68 which contain a more structurally diverse PAH set (Andersson & Achten, 2015; Stout et
69 al., 2015, p. 16). Low molecular weight PAHs are susceptible to weathering, primarily by
70 volatilization, whereas high molecular weight aromatics are more resilient (John, Han &
71 Clement, 2016). Therefore, these are useful targets for oil-oil and oil-source correlation
72 and spill identification and environmental monitoring (Pampanin & Sydnes, 2017;
73 Poulsen et al., 2018).
74
75 Comprehensive identification of the aromatics is challenging due to the large
76 concentration variance. Traditionally, petroleum analysis is based on pre-fractionation
77 using silica chromatography or solid-phase extraction (SPE) cartridges followed by GC-
78 MS^n (Wang, Fingas & Li, 1994; Alzaga et al., 2004; Pillai et al., 2005; Gilgenast et al.,

79 2011). SPE is a low-efficiency separation technique, depending on chemical selectivity.
80 This allows crude isolation of the aromatics fraction, but not separation of the
81 compounds within it (Figure 1). Thus, an aromatic fraction obtained by SPE contains
82 both the benzenes, naphthalenes and larger rings. Here, a typical crude oil will have a
83 high abundance of monoaromatics, with diminishing concentrations with increasing ring
84 size. The appropriate GC on-column concentration of the naphthalenes typically results
85 in the larger ring systems being below the limit of detection (LOD). By increasing
86 concentration to push trace aromatics above the LOD, both the column and detector will
87 be saturated by the more abundant compounds. This leads to high background levels
88 which affect quantitation and may obscure mass spectra complicating structural
89 identification of unknowns (Zhao et al., 2014; Wilton, Wise & Robbat, 2017).
90 Furthermore, the poor resolution of SPE often leads to an overlap between the
91 saturated and aromatic hydrocarbons, which interferes with subsequent analysis. Thus,
92 although SPE is efficient for routine applications, a large portion of the sample remains
93 undetected. Comprehensive multi-dimensional chromatography (GC×GC) is often used
94 as an alternative to simplify or remove the need for sample pre-fractionation (Nizio,
95 McGinitie & Harynuk, 2012; Jennerwein et al., 2014; Stilo et al., 2021). However, it does
96 not solve the issue with variable abundance and column/detector overload. Thus, a
97 complete qualitative and explorative oil analysis requires a more selective sample-
98 prefractionation method.
99
100 Herein, we present a high-performance liquid-chromatography (HPLC) method for the
101 automated high-resolution fractionation of crude oil using commercially available
102 columns. The method can resolve aromatics based on ring size and connectivity, i.e.
103 fused and non-fused rings (*e.g. naphthalene versus biphenyl*). The fractions may be
104 diluted or concentrated, depending on the target, for subsequent analysis and can thus
105 be used to concentrate trace species. The method is easily modified to selectively
106 collect only fractions of interest, and the aromatics may be collected either as one or
107 several fractions. We demonstrate the method's applicability by fractionation of fourteen
108 crude oils with subsequent GC-MS analysis. The method is compared to data obtained
109 using SPE and GC×GC-MS. We demonstrate how it's especially suitable for the
110 analysis and identification of trace aromatics by the successful tentative identification of
111 several compounds not observed using comparable methods.
112

113 ## Materials & Methods
114 ### Chemicals and reagents
115 Chloroform, dichloromethane, n-hexane, deuterated standards and model compounds
116 (ethylbenzene, naphthalene, biphenyl, phenanthrene, 1-benzylnaphthalene and
117 chrysene) were purchased from Sigma Aldrich and used as received.
118

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.

Sample preparation

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.

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.

159 **Analytical methods**

160 *GC-MS*

161 GC-MS data were recorded using an Agilent 5977B GC-MSD as follows; 250 °C inlet,
162 320 °C transfer line, splitless injection (1 μL), Agilent DB-5MS (30 m, 0.25 mm i.d., 0.25
163 μm). The oven temperature gradient was programmed as follows; 50 (1 min. hold-time)
164 - 320 °C (8 min hold-time, 10 °C/min.), helium carrier gas at 1.5 mL/min. in constant flow
165 mode.

166

167 GC×GC-MS data were recorded using an Agilent 7200B GC-QTOF equipped with a
168 Zoex ZX-2 thermal modulator (Zoex Corporation, Houston, TX, USA) as follows; 250 °C
169 inlet, 320 °C transfer line, splitless injection (1 μL), Agilent DB-5MS UI (1D, 30 m, 0.25
170 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)
171 capillary columns connected using a SilTite μ-union. The oven was temperature
172 programmed as follows; 50 (1 min hold-time) - 320 °C (3 °C/min.), helium carrier gas at
173 1 mL/min. in constant flow mode. The modulation period was set to 6 s with a 400 ms
174 hot-jet duration.

175

176 *Data processing*

177 Data were screened using Masshunter Qualitative Navigator (Agilent, B.08.00). Peak
178 detection and compound identification were performed using MassHunter Unknowns
179 Analysis (Agilent, B.09.00) and the NIST Standard Reference Database (1A v17,
180 Gaithersburg, MD, USA). Feature tables were exported as CSV files and imported into a
181 Jupyter Notebook for further processing using the Python scripting language. Duplicates
182 based on the CAS number were removed from the feature tables. All compounds
183 containing silica and halogens were removed. The double-bond equivalent values were
184 calculated for each compound and all features with a DBE of less than 4 were excluded.
185 Finally, experimental and literature retention indices (RI) were compared with flagging of
186 all compounds where the difference was larger than 50 units.

187

188 # Results and Discussion

189 **Method development**

190 The objective of the method was to 1) separate saturates and aromatics and 2) intra-
191 class separation of the aromatics with enrichment capabilities. A dual-column setup
192 using normal phase analytical LC-columns provided the required selectivity and
193 efficiency. The primary column (*Thermo Scientific Hypersil Gold NH₂*) acted as a
194 retainer for polar components, whereas a secondary pure silica-based column (*Thermo*
195 *Scientific Hypersil Silica*) was required for the separation of saturated and aromatic
196 hydrocarbons. The separation was optimized using six model compounds commonly
197 found in crude oil (Figure 2). An isocratic *n*-hexane elution yielded separation of

198 saturated and mono-aromatic hydrocarbons, as well as separation of polycyclic
199 aromatics based on ring size and connectivity (Figure 3).
200
201 The fraction collector was programmed to collect 30 second wide fractions based on the
202 peak widths of the model compounds. At this fraction width, we observed only a minor
203 overlap of fractions with co-elution of the most abundant components. A reduction of the
204 fraction width can be set if a higher peak purity is required. The cost is a slight loss of
205 recovery. After elution of the last aromatics as observed by UV/Vis, the flow path was
206 selectively reversed for the primary column. The column was then rinsed using a
207 gradient from 100% n-hexane to 50:50 chloroform:2-propanol. This effectively removed
208 the adsorbed resins on the amide column. The fraction collector is within the flow path
209 during all stages of chromatography and the resins may therefore be isolated for further
210 analysis (Sundberg & Feilberg, 2020). The final step is a re-equilibration of the whole
211 system by a return to isocratic n-hexane and flushing at an increased flow rate to
212 remove the polar solvents from the flow path. Improper re-equilibration resulted in a
213 severe loss of retention in subsequent fractionations due to the adsorption of 2-
214 isopropanol on the silica phase.
215
216 Recovery values were calculated using the model compounds by comparison of peak
217 areas obtained on GC-MS from LC-fractions compared to direct analysis of the
218 standards (Table 1). Three analytes have recovery values slightly above 100%. This is
219 likely due to a discrepancy between programmed and real injection volume on the
220 HPLC auto-sampler. In contrast, the recovery is less than 90% for chrysene. For this
221 compound, we observe peak broadening due to the high capacity factor. As the fraction
222 collection width is static during the full run, the low recovery is attributed to the peak
223 being wider than the collection width. To evaluate the reproducibility of complex
224 samples, a single oil was fractionated three consecutive times. Each fraction was
225 analyzed on GC-MS and the relative standard deviation was determined from peak
226 areas. 1,2,4-Trimethylbenzene, naphthalene and phenanthrene gave 4.6,8.1 and 6.4%
227 respectively. The results are similar to those obtained using a model mixture. This
228 shows that the method performs consistently in the presence of a highly complex oil
229 matrix. Furthermore, the method successfully removes interferences and yields a high
230 signal-to-noise ratio for the target analytes in each fraction (Figure 4).
231
232 **Applicability in crude oil analyses**
233 The applicability of the method was demonstrated by fractionation and analysis of 14
234 crude oils. The oil samples were sourced from producing fields in the Danish region of
235 the North Sea. Crude oils from this region typically have an aromatics content of 25 –
236 30%, of which the majority are BTEX-type monoaromatics (*benzene, toluene,*
237 *ethylbenzene, xylene*) with a continuous decrease in abundance with increasing ring
238 size (Sundberg & Feilberg, 2020). The primary fraction is the saturated hydrocarbons

239   followed by the resins (up to 5%) with only traces of asphaltenes. This is evident from
240   the fractionation, where a typical dilution factor of 50/20 had to be applied to the
241   saturated and monoaromatic fractions respectively (Table 2). The fractions containing
242   larger aromatics were analyzed either undiluted or concentrated by solvent reduction.
243   The first fraction contains the paraffins and naphthenes and is poorly retained on the
244   primary LC-column (Figure 5). The second silica column is required to separate them
245   from the monoaromatics, which elute as the second fraction (Figure 6). The third and
246   fourth fractions contain diaromatic species, with the latter non-fused ring systems (e.g.
247   naphthalene versus biphenyl). Fractions 5 and 6 contain the triaromatics (e.g.
248   phenanthrene versus 1-phenylnaphthalene). Here, the abundance starts to diminish and
249   the sixth fraction had to be concentrated for subsequent GC-MS analysis. Fractions 7
250   and above contain larger ring systems, e.g. chrysene, perylene. These fractions are
251   less well-defined, likely because compounds eluting within this retention range are
252   fewer in number and present in trace amounts. We also observed a slight loss of
253   resolution, with minor overlap and cross-contamination. This is a result of two things; 1)
254   diffusion and peak broadening during the liquid chromatography 2) collection of low
255   abundance (undiluted/concentrated) fraction after a high abundance (diluted) fraction. If
256   higher purity peaks are required the fraction collection width can be reduced. Attempts
257   to concentrate fractions 9 and later were not successful and the gas chromatograms
258   were dominated by background contaminants likely originating from the solvents,
259   HPLC-tubing and glassware (e.g. siloxanes, surfactants).
260
261   **Performance comparison with SPE-GC-MS and GC×GC-MS**
262   Solid-phase extraction of crude oil into its saturated and aromatic fraction is a well-
263   established sample preparation method. The physical properties of SPE adsorbents
264   (large particle size, low mass loadings) result in limited separation power (Berrueta,
265   Gallo & Vicente, 1995; Buszewski & Szultka, 2012). Thus, the technique is mainly
266   applicable for the crude separation of different compound classes. It does not provide
267   sufficient resolution to separate closely related compounds within subfractions. To
268   compare our method to SPE we fractionated each oil using a Phenomenex Strata EPH
269   (200 μm, 70 Å, 500 mg / 3 mL). The cartridge contains a proprietary phase specifically
270   developed to separate hydrocarbon fractions (Countryman, Kelly & Garriques, 2005).
271
272   In terms of spectral quality, an approximately 10-fold reduction in background noise is
273   observed in the LC fractions as compared to SPE. A comparison of the extracted mass
274   spectra for the peak corresponding to 1,3-dimethylpyrene is presented in Figure 7.
275   Selected ion monitoring can be used to reduce background interferences for target
276   species but results in loss of spectral detail for qualitative analysis. Furthermore, when
277   using low-resolution instruments, i.e. single quadrupole MS, there is a large risk of
278   overlap in complex samples (Rosenthal, 1982; Davis & Giddings, 1983). The reduction

279 in background noise improved library matching, especially for analytes present at trace
280 levels.
281
282 To evaluate identification performance, peak picking and library matching were carried
283 out using MassHunter Unknowns Analysis and the NIST mass spectral library. The
284 match factor limit was set to 700. The number of compounds was compared both on a
285 sample-to-sample basis and by merging all features from all samples (with duplicates
286 removal based on CAS number). A comparison of the merged compound tables of all
287 samples shows that using the LC-GC method we can identify 957 compared to 601
288 compounds using SPE. This is an increase of 37.2%. To increase the match
289 confidence, we applied a retention index (RI) filter, only retaining compounds with a
290 match within 100 units of the library value. By doing so, we identified 426 compared to
291 300 (42% increase). This excludes all compounds of which a library RI is not available
292 (approximately 1% of our feature set). However, a large portion of the compounds only
293 have computationally approximated retention indices and not experimentally determined
294 values. Thus, all filtering and data analyses should be carried out with care and manual
295 intervention. The SPE fraction contains approximately 190 unique compounds with 404
296 compounds overlapping both analyses. Manual inspection reveals this list contains
297 several petroleum-type compounds and not predominantly background noise or
298 contamination (e.g. plasticizers, column contamination). One plausible source is errors
299 occurring during the automatic processing routines. Small differences in mass spectra
300 (e.g. due to abundance or background level) can lead to closely related library matches
301 being given similar (but different) priority (e.g. isomeric species). Figure 8 shows the
302 DBE distributions of assigned compounds uniquely observed from SPE-GC and LC-GC
303 methods. Noticeably, the DBE distributions are significantly different between the two
304 methods. The distribution of unique compounds from SPE-GC is centered around low
305 DBE 4 and 5 (e.g. monoaromatic), whereas unique compounds from LC-GC are
306 distributed more evenly at higher DBE values. This is expected as the LC-GC method
307 isolates and enriches high aromaticity fractions. These findings showcase the ability of
308 LC-GC as a high-resolution fractionation method for crude oil.
309
310 For comparison to comprehensive multi-dimensional chromatography, the samples
311 were analyzed by our in-house routine GC×GC-MS method (i.e. solvent dilution,
312 filtration and analysis) (Figure 9). The objective of the GC×GC method is not to
313 maximize feature ID but enable multi-class analysis/fingerprinting with minimal to no
314 sample preparation. Furthermore, the SPE-LC-GC and GC×GC analyses were carried
315 out on different instruments which makes direct comparison challenging. For GC×GC an
316 Agilent 7200B QTOF high-resolution mass spectrometer was used. For SPE-LC-GC, an
317 Agilent 5977B single quadrupole equipped with a High-Efficiency Source (HES) was
318 used. The HES has both higher sensitivity and dynamic range. Secondly, for GC×GC,

319 the dilution factor was adjusted so that the analytes with the highest abundance were at
320 detector saturation. Here, we see that although GC×GC is not restricted in terms of
321 peak capacity, it does fall short in terms of dynamic range. After blob detection, library
322 matching and filtering we obtain 63 tentative hits in a single sample. With corresponding
323 processing settings, we identified 143 compounds by multi-fraction LC-GC-MS analysis
324 of the same sample. This is an increase of 127%. In terms of manual intervention, ease
325 of use and time of analysis, GC×GC is preferred compared to LC-GC-MS. However, the
326 amount of data generated using the latter is more comprehensive in our case.
327
328 A Venn diagram was constructed to compare three methods (Figure 8). The number
329 within each colored circle represents the number of assigned unique compounds for
330 each method, whereas numbers in overlapped zones represent the number of
331 compounds that have been co-assigned from corresponding methods. The amount of
332 compositions obtained from LCxGC (957) significantly surpasses GC×GC (181) and
333 SPE-GC (601). We obtained approximately 50% unique compounds with GC×GC and
334 LC-GC and 32% with SPE-GC. The Venn diagram also shows that co-assigned
335 compounds of those three methods cover a narrow range of overall chemical
336 composition (59 co-assigned compounds) of crude oil. Again, it is worth noting that
337 there are differences in terms of dilution factor and instrumental parameters for those
338 methods. Therefore, the comparison is biased but still relevant to evaluate the LC-GC
339 method for trace components analysis.
340

341 **Conclusions**
342 We have developed a method for high-resolution fractionation of complex crude oil
343 matrices. By using sub-micron LC columns we obtained high efficiency and resolution
344 which allowed intra-class compound separation. This is in contrast with traditional
345 methods, e.g. SPE, which yields a single aromatics fraction. The method is especially
346 advantageous for the isolation of trace species. Multiple compounds not observed by
347 SPE-GC-MS were pre-concentrated yielding high abundance and spectral quality. The
348 increase in the number of tentatively identified peaks is thus a result of both reduced co-
349 elution and an increase in analyte signal-to-noise ratio.
350

351 By characterization of 14 crude oils, we extended the identification to a large number of
352 hydrocarbon and N,S,O-containing aromatics. Of the 517 uniquely identified
353 compounds, 69% (357) contain either N,S,O (or a combination of) atoms (Table 3). The
354 structures of five representative compounds are presented in Figure 10. Aromatic
355 nitrogen and sulfur compounds are detrimental in petroleum processing. Furthermore,
356 they potentially have biological activity and may pose an environmental and
357 toxicological hazard (López García et al., 2002; Anyanwu & Semple, 2015; Zhang et al.,
358 2018; Vetere, Pröfrock & Schrader, 2021). Therefore, their characterization is an

359 important pursuit. They are routinely analyzed by direct infusion mass spectrometry that
360 provides the molecular formula but not connectivity (Guan, Marshall & Scheppele, 1996;
361 Purcell et al., 2007b,a; Corilo, Rowland & Rodgers, 2016). Thus, isolation and GC-MS
362 analysis with library matching provide valuable information on their presence in oil
363 samples.
364
365 The relatively long fractionation time (60 minutes) and the number of fractions
366 generated lead to a full sample analysis time of 6 hours (when characterizing the first 7
367 fractions using GC-MS). Several steps require manual intervention, i.e. dilution and pre-
368 concentration of fractions and moving the samples from the LC to the GC. It would
369 therefore be beneficial to implement more automation, e.g. by using liquid handling
370 robotics (ultimately with direct hyphenation to the GC). We observed minor co-elution
371 during the analysis of latter fractions. Combining the LC-fractionation with subsequent
372 GC×GC analysis would increase the power of the method further. However, it would
373 require an intense data processing workflow with high demands in computational power.
374 Something that is already challenging in comprehensive GC×GC studies (Reichenbach
375 et al., 2019; Wilde et al., 2020; Stefanuto, Smolinska & Focant, 2021).
376

Acknowledgments

378 The authors are grateful for the donation of the samples from Total E&P Denmark.
379 Furthermore, the authors wish to thank Karen L. Feilberg for discussions.

380

References

382 Aftalion F. 2001. *A history of the international chemical industry*. Philadelphia, PA:

383   Chemical Heritage Press.

384 Alzaga R, Montuori P, Ortiz L, Bayona JM, Albaigés J. 2004. Fast solid-phase

385   extraction–gas chromatography–mass spectrometry procedure for oil

386   fingerprinting. *Journal of Chromatography A* 1025:133–138. DOI:

387   10.1016/j.chroma.2003.10.080.

388 Andersson JT, Achten C. 2015. Time to Say Goodbye to the 16 EPA PAHs? Toward an

389   Up-to-Date Use of PACs for Environmental Purposes. *Polycyclic Aromatic*

390   *Compounds* 35:330–354. DOI: 10.1080/10406638.2014.991042.

391   Anyanwu IN, Semple KT. 2015. Fate and behaviour of nitrogen-containing polycyclic

392   aromatic hydrocarbons in soil. *Environmental Technology & Innovation* 3:108–

393   120. DOI: 10.1016/j.eti.2015.02.006.

394   Berrueta LA, Gallo B, Vicente F. 1995. A review of solid phase extraction: Basic

395   principles and new developments. *Chromatographia* 40:474–483. DOI:

396   10.1007/BF02269916.

397   Buszewski B, Szultka M. 2012. Past, Present, and Future of Solid Phase Extraction: A

398   Review. *Critical Reviews in Analytical Chemistry* 42:198–213. DOI:

399   10.1080/07373937.2011.645413.

400   Carvalho Dias L, Bomfim Bahia PV, Nery do Amaral D, Machado ME. 2020. Nitrogen

401   compounds as molecular markers: An overview of analytical methodologies for

402   its determination in crude oils and source rock extracts. *Microchemical Journal*

403   157:105039. DOI: 10.1016/j.microc.2020.105039.

404   Corilo YE, Rowland SM, Rodgers RP. 2016. Calculation of the Total Sulfur Content in

405   Crude Oils by Positive-Ion Atmospheric Pressure Photoionization Fourier

406   Transform Ion Cyclotron Resonance Mass Spectrometry. *Energy & Fuels*

407   30:3962–3966. DOI: 10.1021/acs.energyfuels.6b00497.

408   Countryman S, Kelly K, Garriques M. 2005. An Improved Method for Diesel

409   Fractionation Following the Revised Massachusetts EPH Guidelines.

410   Davis JM, Giddings JC. 1983. Statistical theory of component overlap in

411   multicomponent chromatograms. *Analytical Chemistry* 55:418–424. DOI:

412   10.1021/ac00254a003.

413   Gilgenast E, Boczkaj G, Przyjazny A, Kamiński M. 2011. Sample preparation procedure

414   for the determination of polycyclic aromatic hydrocarbons in petroleum vacuum

415   residue and bitumen. *Analytical and Bioanalytical Chemistry* 401:1059–1069.

416   DOI: 10.1007/s00216-011-5134-9.

417   Guan S, Marshall AG, Scheppele SE. 1996. Resolution and Chemical Formula

418   Identification of Aromatic Hydrocarbons and Aromatic Compounds Containing

419   Sulfur, Nitrogen, or Oxygen in Petroleum Distillates and Refinery Streams.

420   *Analytical Chemistry* 68:46–71. DOI: 10.1021/ac9507855.

421   Jennerwein MK, Eschner M, Gröger T, Wilharm T, Zimmermann R. 2014. Complete

422   Group-Type Quantification of Petroleum Middle Distillates Based on

423   Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass

424   Spectrometry (GC×GC-TOFMS) and Visual Basic Scripting. *Energy & Fuels*

425   28:5670–5681. DOI: 10.1021/ef501247h.

426   John GF, Han Y, Clement TP. 2016. Weathering patterns of polycyclic aromatic

427   hydrocarbons contained in submerged Deepwater Horizon oil spill residues when

428   re-exposed to sunlight. *Science of The Total Environment* 573:189–202. DOI:

429   10.1016/j.scitotenv.2016.08.059.

430   Keith LH. 2015. The Source of U.S. EPA's Sixteen PAH Priority Pollutants. *Polycyclic*

431   *Aromatic Compounds* 35:147–160. DOI: 10.1080/10406638.2014.892886.

432   Krajewski LC, Rodgers RP, Marshall AG. 2017. 126 264 Assigned Chemical Formulas

433   from an Atmospheric Pressure Photoionization 9.4 T Fourier Transform Positive

434   Ion Cyclotron Resonance Mass Spectrum. *Analytical Chemistry* 89:11318–

435   11324. DOI: 10.1021/acs.analchem.7b02004.

436 Lawal AT. 2017. Polycyclic aromatic hydrocarbons. A review. *Cogent Environmental*

437   *Science* 3:1339841. DOI: 10.1080/23311843.2017.1339841.

438 López García C, Becchi M, Grenier-Loustalot MF, Païsse O, Szymanski R. 2002.

439   Analysis of Aromatic Sulfur Compounds in Gas Oils Using GC with Sulfur

440   Chemiluminescence Detection and High-Resolution MS. *Analytical Chemistry*

441   74:3849–3857. DOI: 10.1021/ac011190e.

442 Marshall AG, Rodgers RP. 2008. Petroleomics: Chemistry of the underworld.

443   *Proceedings of the National Academy of Sciences* 105:18090–18095. DOI:

444   10.1073/pnas.0805069105.

445 Mössner SG, Wise SA. 1999. Determination of Polycyclic Aromatic Sulfur Heterocycles

446   in Fossil Fuel-Related Samples. *Analytical Chemistry* 71:58–69. DOI:

447   10.1021/ac980664f.

448 Net Zero by 2050 A Roadmap for the Global Energy Sector. 2021.

449 Nizio KD, McGinitie TM, Harynuk JJ. 2012. Comprehensive multidimensional

450   separations for the analysis of petroleum. *Journal of Chromatography A*

451   1255:12–23. DOI: 10.1016/j.chroma.2012.01.078.

452 Palacio Lozano DC, Gavard R, Arenas-Diaz JP, Thomas MJ, Stranz DD, Mejía-Ospino

453   E, Guzman A, Spencer SEF, Rossell D, Barrow MP. 2019. Pushing the analytical

454   limits: new insights into complex mixtures using mass spectra segments of

455   constant ultrahigh resolving power. *Chemical Science* 10:6966–6978. DOI:

456   10.1039/C9SC02903F.

457 Pampanin DM, Sydnes MO (eds.). 2017. *Petrogenic Polycyclic Aromatic Hydrocarbons*

458   *in the Aquatic Environment: Analysis, Synthesis, Toxicity and Environmental*

459   *Impact*. BENTHAM SCIENCE PUBLISHERS. DOI:

460   10.2174/9781681084275117010l.

461   Pillai I, Ritchie L, Heywood R, Wilson G, Pahlavanpour B, Setford S, Saini S. 2005.

462   Development of an improved analytical method for the determination of

463   carcinogenic polycyclic aromatic hydrocarbons in transformer oil. *Journal of*

464   *Chromatography A* 1064:205–212. DOI: 10.1016/j.chroma.2004.12.066.

465   Poulsen KG, Kristensen M, Tomasi G, Dela Cruz M, Christensen JH. 2018. The Pixel-

466   Based Chemometric Approach for Oil Spill Identification and Hydrocarbon

467   Source Differentiation. In: *Oil Spill Environmental Forensics Case Studies*.

468   Elsevier, 443–463. DOI: 10.1016/B978-0-12-804434-6.00021-5.

469   Purcell JM, Juyal P, Kim D-G, Rodgers RP, Hendrickson CL, Marshall AG. 2007a.

470   Sulfur Speciation in Petroleum: Atmospheric Pressure Photoionization or

471   Chemical Derivatization and Electrospray Ionization Fourier Transform Ion

472   Cyclotron Resonance Mass Spectrometry. *Energy & Fuels* 21:2869–2874. DOI:

473   10.1021/ef700210q.

474   Purcell JM, Rodgers RP, Hendrickson CL, Marshall AG. 2007b. Speciation of nitrogen

475   containing aromatics by atmospheric pressure photoionization or electrospray

476   ionization fourier transform ion cyclotron resonance mass spectrometry. *Journal*

477   *of the American Society for Mass Spectrometry* 18:1265–1273. DOI:

478   10.1016/j.jasms.2007.03.030.

479   Reichenbach SE, Tao Q, Cordero C, Bicchi C. 2019. A Data-Challenge Case Study of

480   Analyte Detection and Identification with Comprehensive Two-Dimensional Gas

481   Chromatography with Mass Spectrometry (GC×GC-MS). *Separations* 6:38. DOI:

482   10.3390/separations6030038.

483 Requejo AG, Sassen R, McDonald T, Denoux G, Kennicutt MC, Brooks JM. 1996.

484   Polynuclear aromatic hydrocarbons (PAH) as indicators of the source and

485   maturity of marine crude oils. *Organic Geochemistry* 24:1017–1033. DOI:

486   10.1016/S0146-6380(96)00079-4.

487 Rosenthal David. 1982. Theoretical limitations of gas chromatographic/mass

488   spectrometric identification of multicomponent mixtures. *Analytical Chemistry*

489   54:63–66. DOI: 10.1021/ac00238a020.

490 Stefanuto P-H, Smolinska A, Focant J-F. 2021. Advanced chemometric and data

491   handling tools for GC×GC-TOF-MS. *TrAC Trends in Analytical Chemistry*

492   139:116251. DOI: 10.1016/j.trac.2021.116251.

493 Stilo F, Bicchi C, Jimenez-Carvelo AM, Cuadros-Rodriguez L, Reichenbach SE,

494   Cordero C. 2021. Chromatographic fingerprinting by comprehensive two-

495   dimensional chromatography: Fundamentals and tools. *TrAC Trends in Analytical*

496   *Chemistry* 134:116133. DOI: 10.1016/j.trac.2020.116133.

497 Stout SA, Emsbo-Mattingly SD, Douglas GS, Uhler AD, McCarthy KJ. 2015. Beyond 16

498   Priority Pollutant PAHs: A Review of PACs used in Environmental Forensic

499   Chemistry. *Polycyclic Aromatic Compounds* 35:285–315. DOI:

500   10.1080/10406638.2014.891144.

501 Sundberg J, Feilberg KL. 2020. Characterization of heteroatom distributions in the polar

502   fraction of North Sea oils using high-resolution mass spectrometry. *Journal of*

503   *Petroleum Science and Engineering* 184:106563. DOI:

504   10.1016/j.petrol.2019.106563.

505 Vetere A, Pröfrock D, Schrader W. 2021. Qualitative and Quantitative Evaluation of

506   Sulfur-Containing Compound Types in Heavy Crude Oil and Its Fractions. *Energy*

507   *& Fuels* 35:8723–8732. DOI: 10.1021/acs.energyfuels.1c00491.

508 Wang Z, Fingas M, Li K. 1994. Fractionation of a Light Crude Oil and Identification and

509   Quantitation of Aliphatic, Aromatic, and Biomarker Compounds by GC-FID and

510   GC-MS, Part I. *Journal of Chromatographic Science* 32:361–366. DOI:

511   10.1093/chromsci/32.9.361.

512 Wei O, Xu X, Zhang Y, Yang B, Ye Q, Yang Z. 2018. Multidimensional Gas

513   Chromatography–Mass Spectrometry Method for Fingerprinting Polycyclic

514   Aromatic Hydrocarbons and Their Alkyl-Homologs in Crude Oil. *Analytical Letters*

515   51:483–495. DOI: 10.1080/00032719.2017.1341905.

516 Wilde MJ, Zhao B, Cordell RL, Ibrahim W, Singapuri A, Greening NJ, Brightling CE,

517   Siddiqui S, Monks PS, Free RC. 2020. Automating and Extending

518   Comprehensive Two-Dimensional Gas Chromatography Data Processing by

519   Interfacing Open-Source and Commercial Software. *Analytical Chemistry*

520   92:13953–13960. DOI: 10.1021/acs.analchem.0c02844.

521 Wilton NM, Wise SA, Robbat A. 2017. Errors in alkylated polycyclic aromatic

522   hydrocarbon and sulfur heterocycle concentrations caused by currently employed

523   standardized methods. *Analytica Chimica Acta* 977:20–27. DOI:

524   10.1016/j.aca.2017.04.017.

525   Yadav VG, Yadav GD, Patankar SC. 2020. The production of fuels and chemicals in the

526   new world: critical analysis of the choice between crude oil and biomass vis-à-vis

527   sustainability and the environment. *Clean Technologies and Environmental*

528   *Policy* 22:1757–1774. DOI: 10.1007/s10098-020-01945-5.

529   Zhang G, Yang C, Serhan M, Koivu G, Yang Z, Hollebone B, Lambert P, Brown CE.

530   2018. Characterization of Nitrogen-Containing Polycyclic Aromatic Heterocycles

531   in Crude Oils and Refined Petroleum Products. In: *Advances in Marine Biology*.

532   Elsevier, 59–96. DOI: 10.1016/bs.amb.2018.09.006.

533   Zhao Y, Hong B, Fan Y, Wen M, Han X. 2014. Accurate analysis of polycyclic aromatic

534   hydrocarbons (PAHs) and alkylated PAHs homologs in crude oil for improving

535   the gas chromatography/mass spectrometry performance. *Ecotoxicology and*

536   *Environmental Safety* 100:242–250. DOI: 10.1016/j.ecoenv.2013.10.018.

537

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.

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%)	

Chemistry Journals
Analytical, Inorganic, Organic, Physical, Materials Science

Summary description of the major constituents of each fraction.

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.

Chemistry Journals

Analytical, Inorganic, Organic, Physical, Materials Science

Comparison of the number of tentatively identified species in SPE-GC and GC×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.

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

Figure 1

Schematic of the three sample preparation and analytical strategies discussed in the paper showing the extent of fractionation possible.

Grey boxes indicate the individual samples that are characterized by mass spectrometry.

11111111111111111111111111111111111111111111111

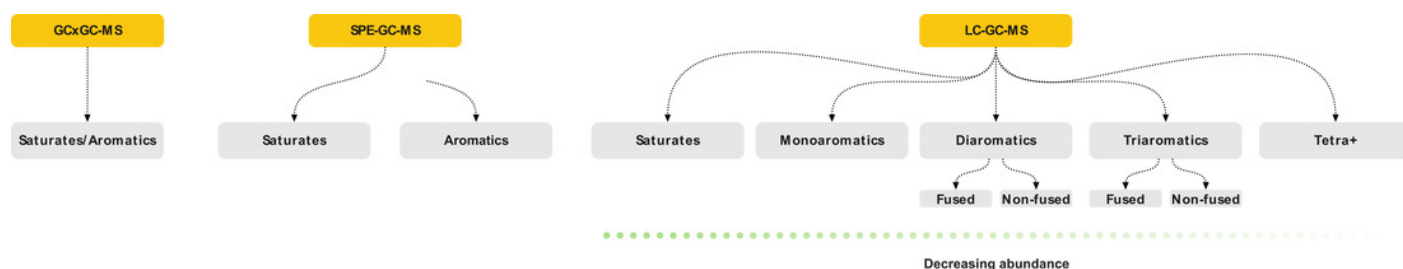



Figure 2

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

![PeerJ] **Chemistry Journals**
Analytical, Inorganic, Organic, Physical, Materials Science

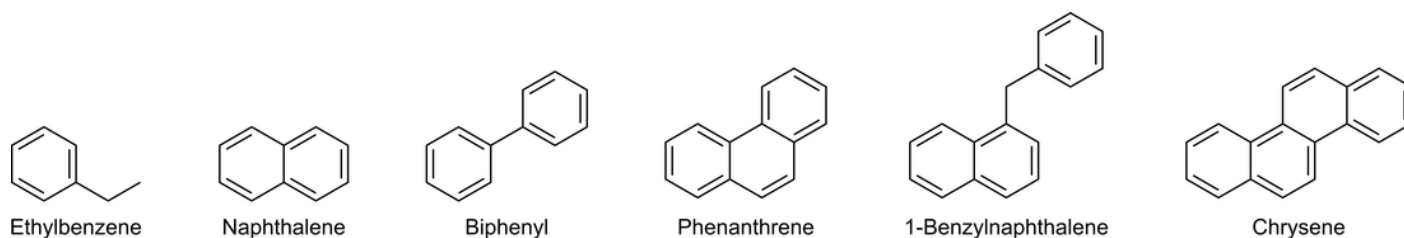



Figure 3

HPLC-Chromatogram showing the elution order and separation of the model compounds used for method development.

![PeerJ] **Chemistry Journals**
Analytical, Inorganic, Organic, Physical, Materials Science

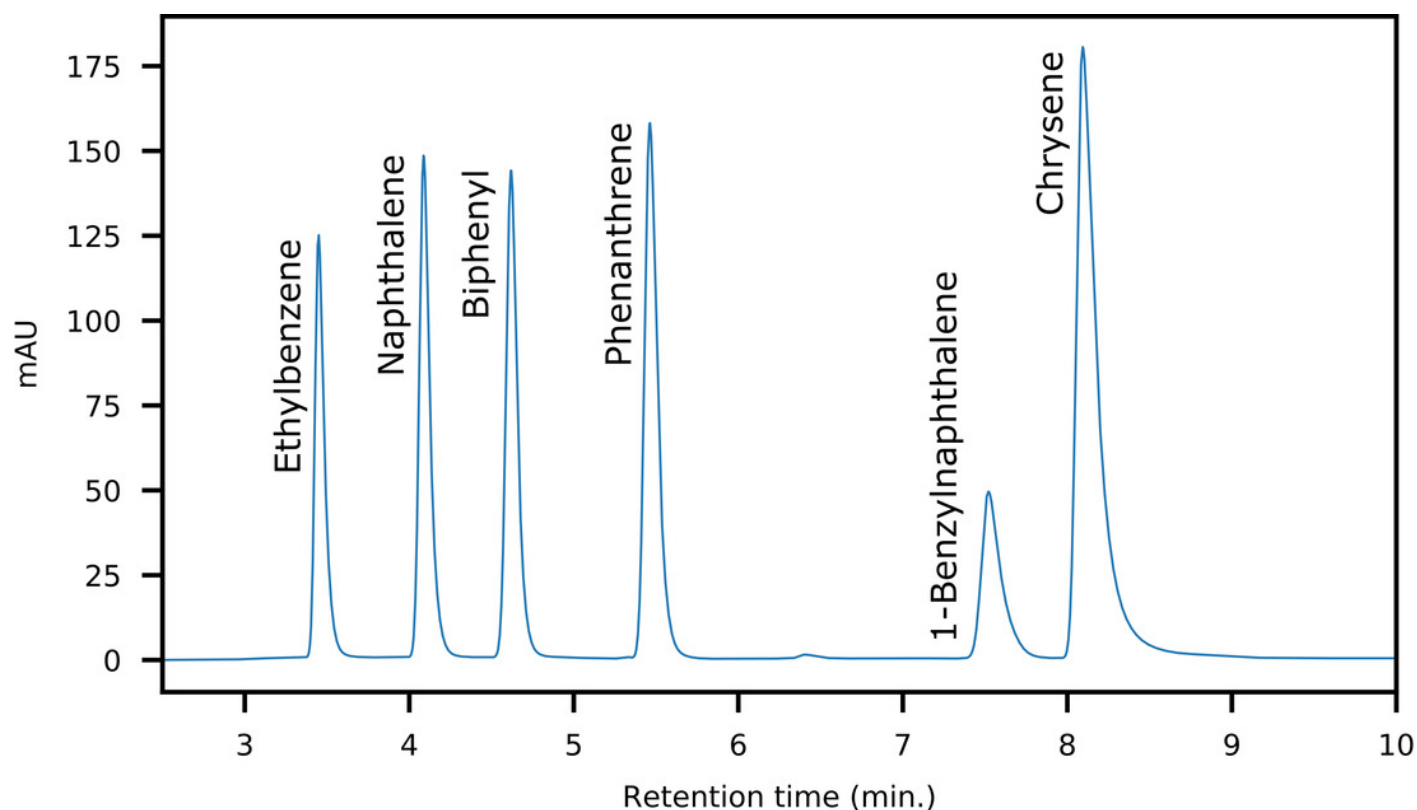



Figure 4

Chromatogram showing the peak corresponding to phenanthrene obtained via SPE-extractions (A) as compared to LC-fractionation (B).

The sample was fractionated using each method in triplicate to determine reproducibility.

Signal intensity variance between A and B is due to different dilution factors (on-column).

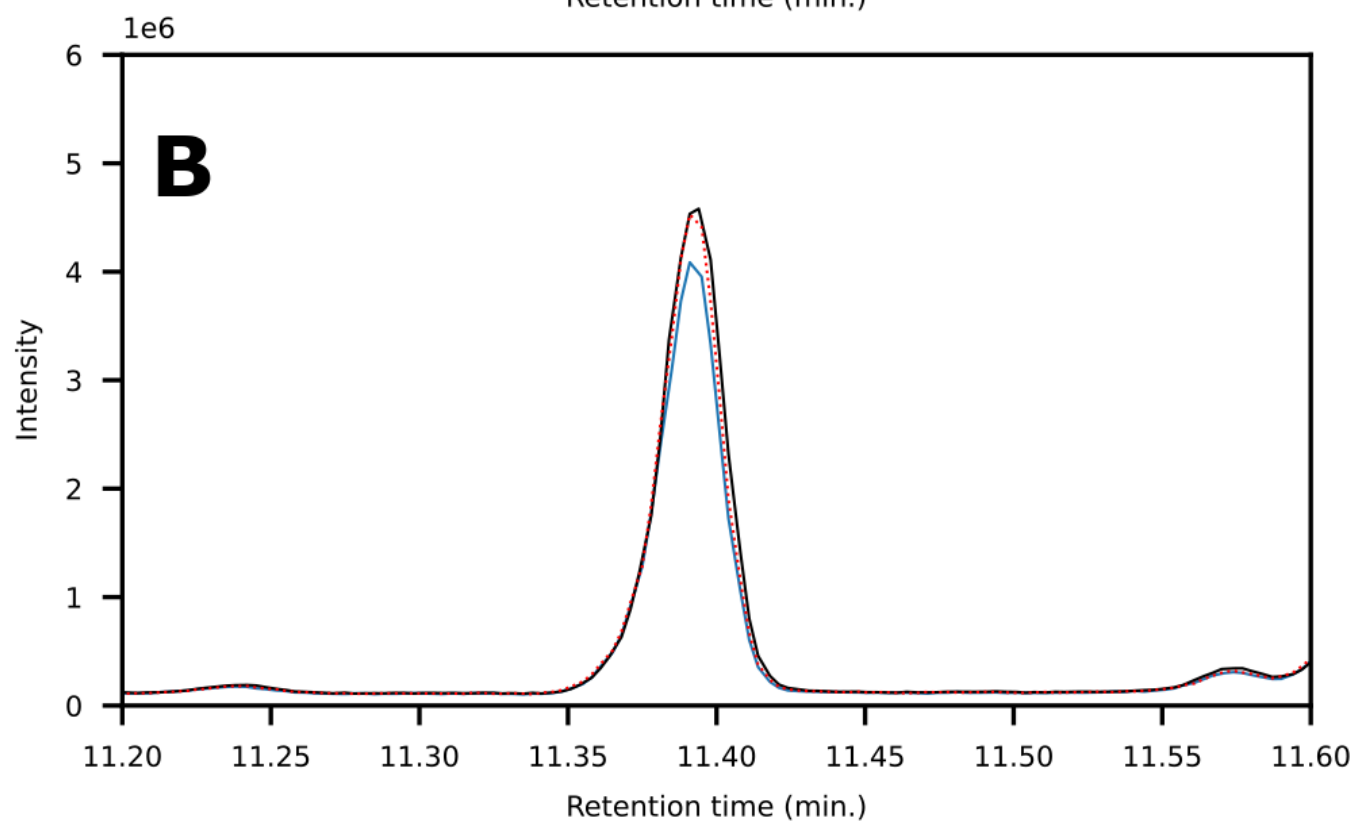
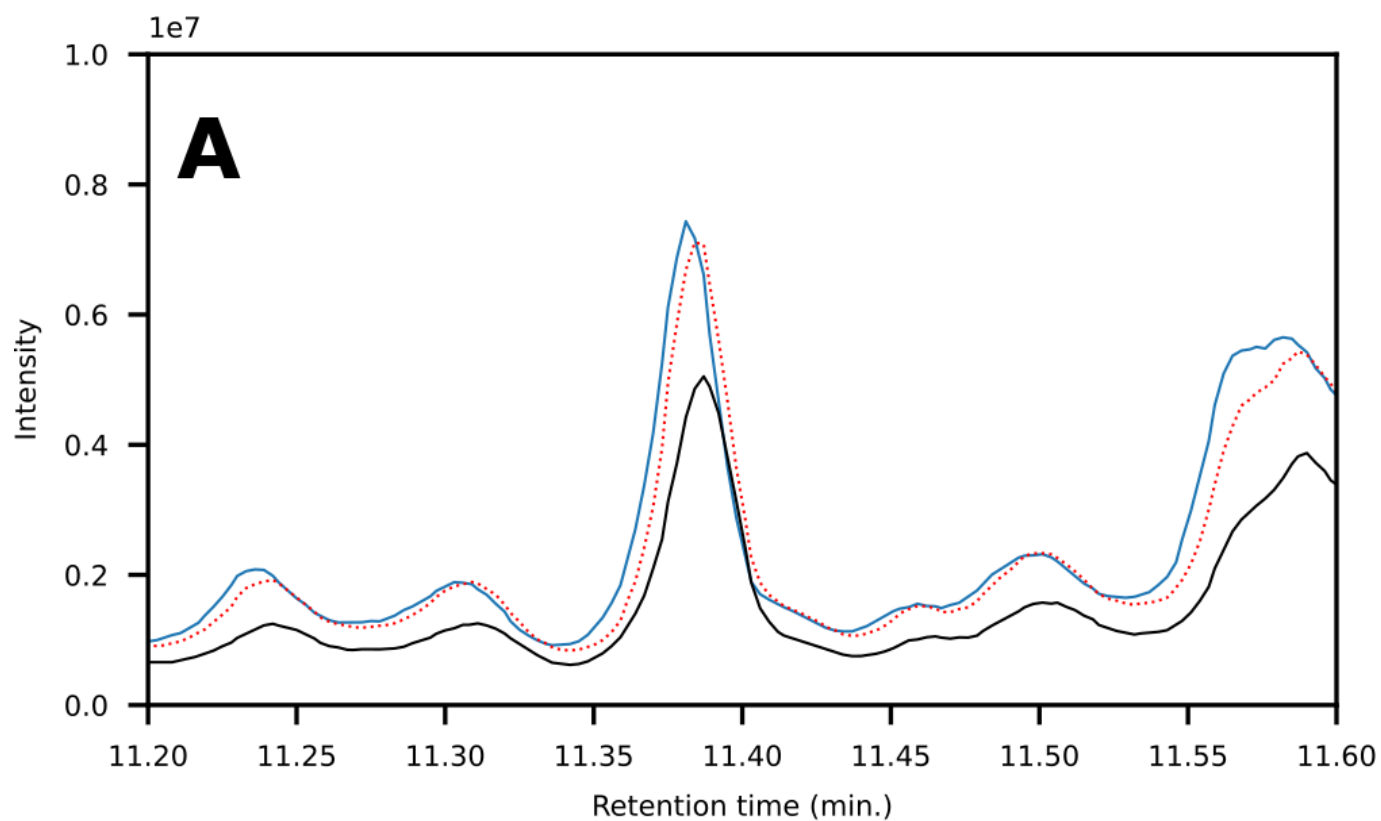



Figure 5

HPLC-Chromatogram (with UV response measured at 272 nm) of a representative crude oil showing the approximate retention regions of the different aromatic classes.

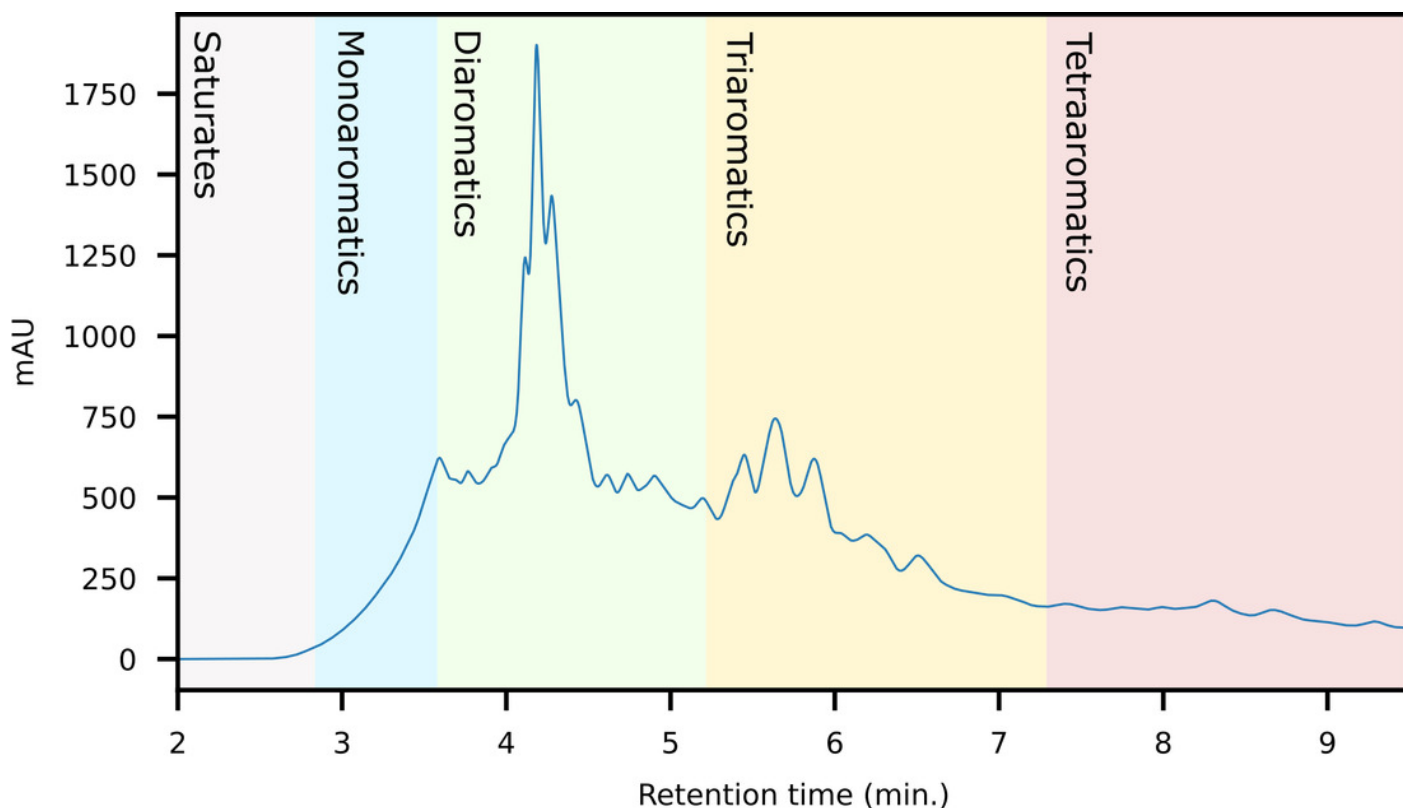



Figure 6

Offset overlay of the GC-MS chromatograms of fractions 2 to 5 obtained using the LC method of a representative sample.

Typical compounds in each fraction have been marked to demonstrate the separating power.

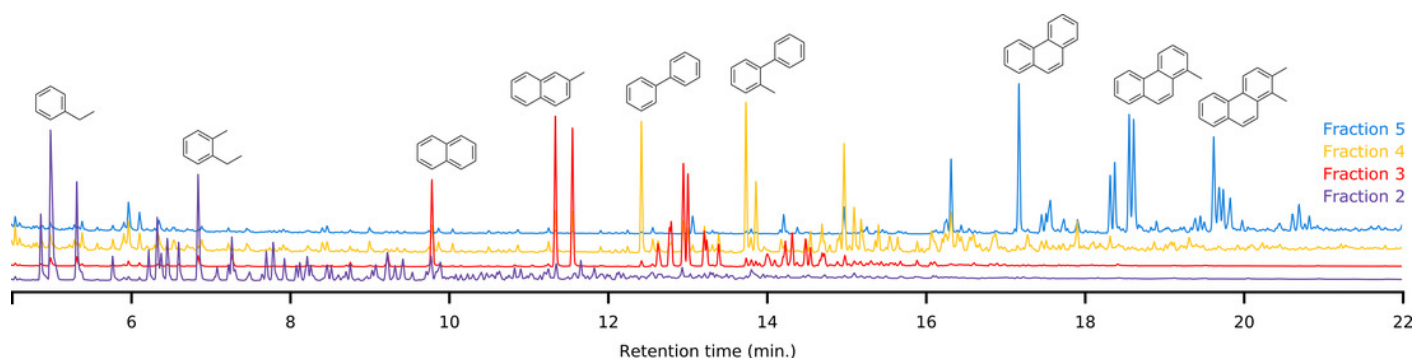



Figure 7

A: Chromatograms showing the peak of 1,3-dimethylpyrene (marked in blue) found in LC-fraction 7 (black, pre-concentrated) and the aromatics fraction obtained via SPE (dotted red). B: Mass spectra mirror plot of the LC-fraction peak showcasing its purity.

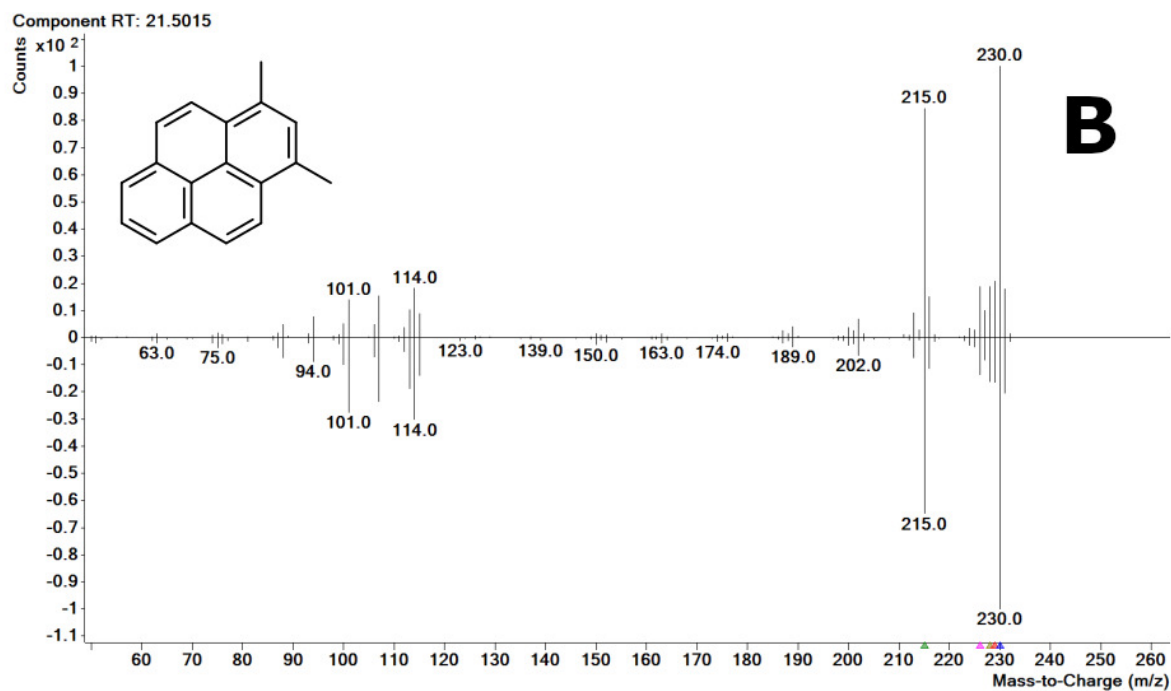
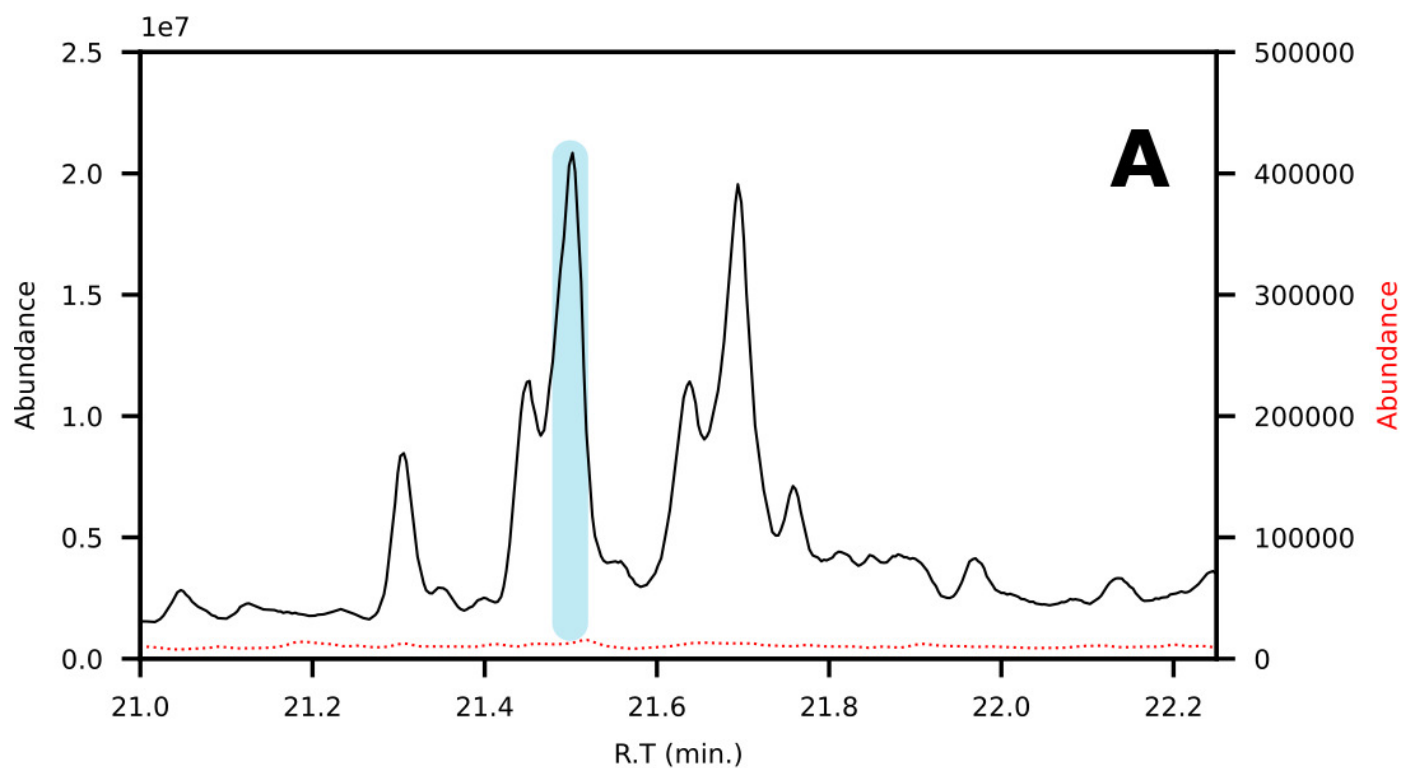



Figure 8

Double bond equivalent (DBE) distribution table for assigned unique compounds obtained from SPE-GC and LC-GC methods. The Venn diagram shows the total amount of assigned compositions obtained from SPE-GC-, LC-GC- and GCxGC-MS.

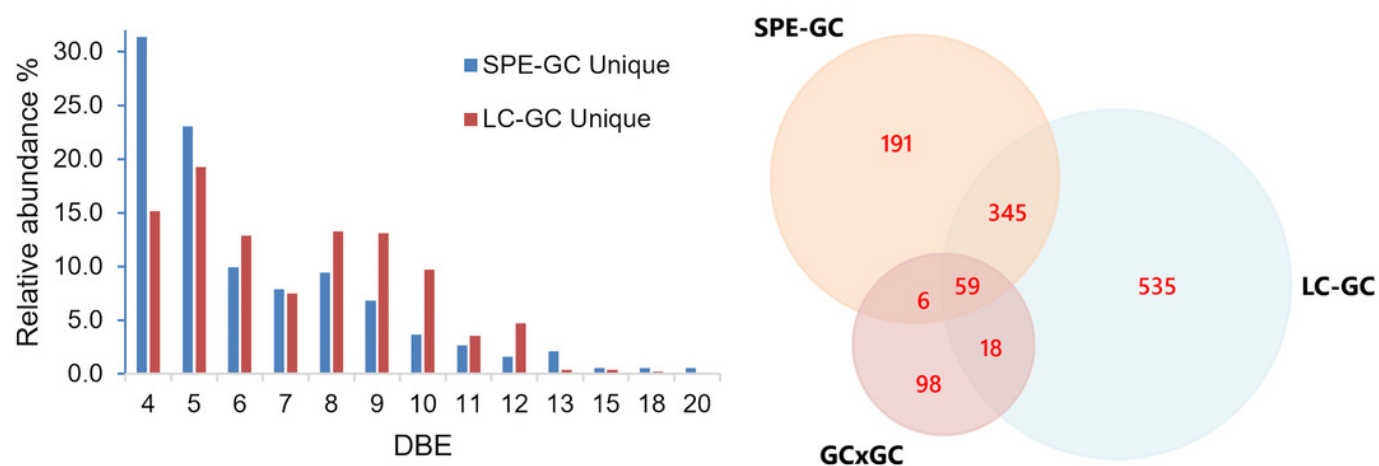



Figure 9

GC×GC-Chromatograms of fraction 5 obtained using the LC-method (A) compared to whole oil analysis (B, i.e. no pre-fractionation).

The LC-fraction has a high abundance of alkyl phenanthrene isomers at or below the LOD seen in whole oil GC×GC-MS.

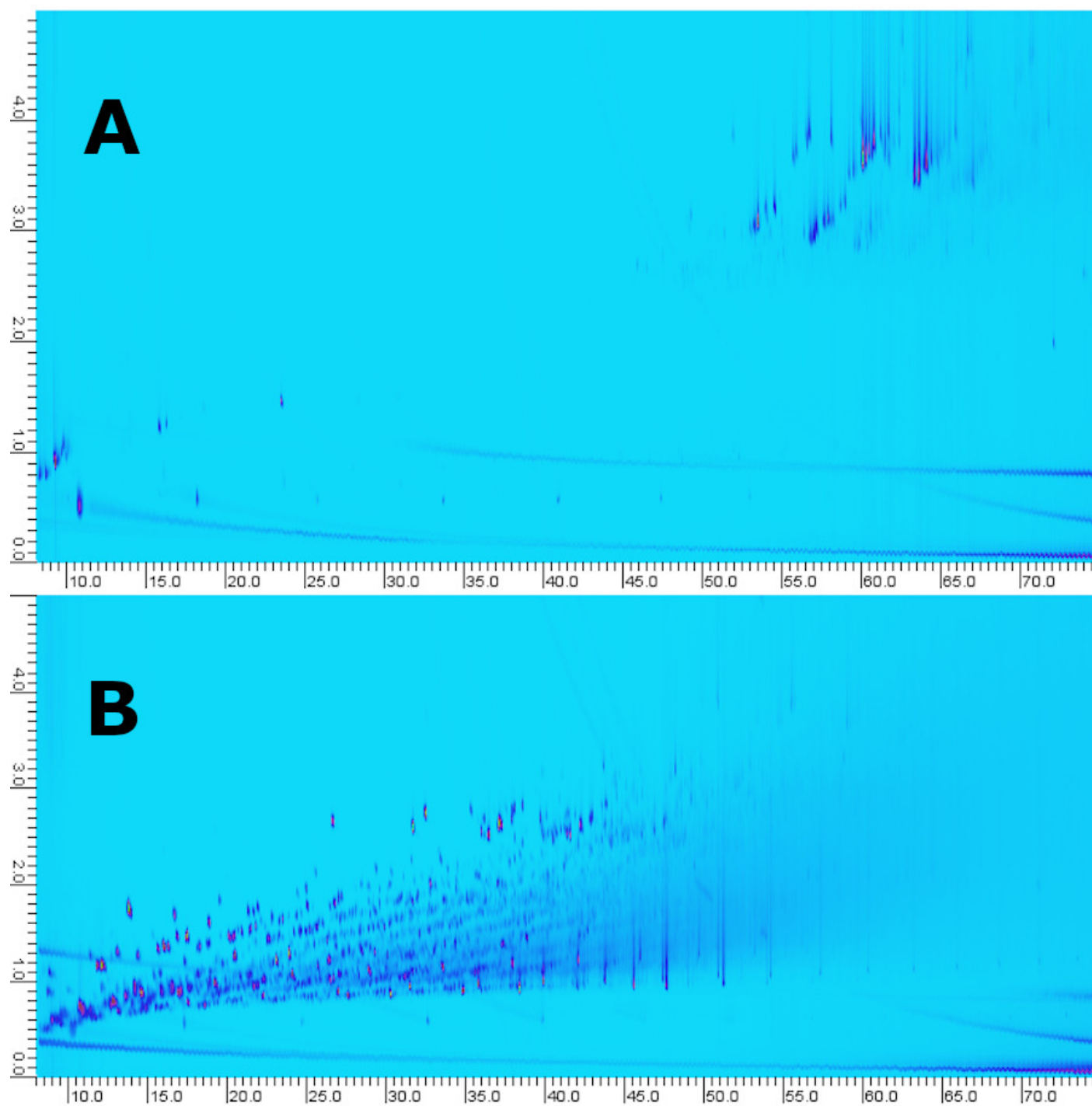
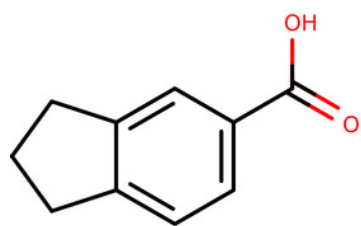
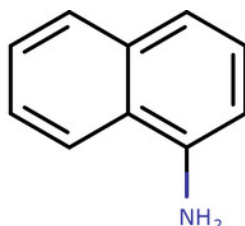
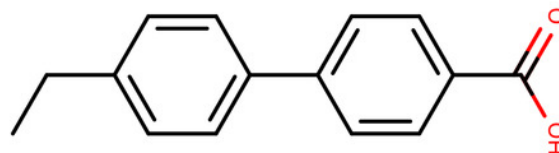
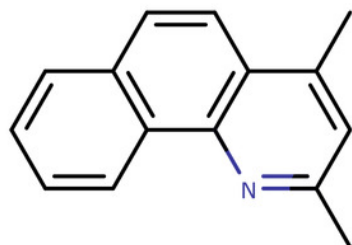
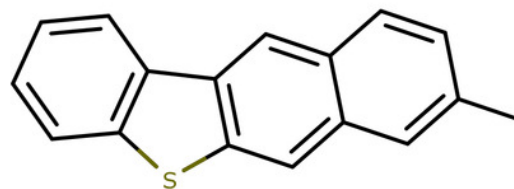



Figure 10

The molecular structures of five unique compounds identified in separate fractions (2 to 6) obtained by using the LC-method.

Indan-5-carboxylic acid
Fraction 2

1-Naphthylamine
Fraction 3

4'-Ethyl-4-biphenylcarboxylic acid
Fraction 4

2,4-Dimethylbenzo[h]quinoline
Fraction 5

8-Methyl-benzo[b]naphtho[2,3-d]thiophene
Fraction 6