The characterization of water-soluble organic compounds released from organic-rich sediments

vorgelegt von
M.Sc.
Jacqueline Mireya Calzada Mendoza
geb. in Mexiko Stadt, Mexiko

von der Fakultät VI – Planen, Bauen, Umwelt
der Technischen Universität Berlin
zur Erlangung des akademischen Grades

Doktor der Naturwissenschaften
- Dr. rer. nat. -

genehmigte Dissertation

Promotionsausschuss:

Vorsitzender: Prof. Dr. Wilhelm Dominik
Gutachter: Prof. Dr. Brian Horsfield
Gutachter: Prof. Dr. Heinz Wilkes

Tag der wissenschaftlichen Aussprache: 19. December 2017

Berlin, 2018

Para mi querida familia

"*En busca de un sueño*
hermoso y rebelde.
En busca de un sueño
que gana y que pierde."
Silvio Rodríguez

Abstract

The study of deep biosphere, especially the growth under extreme conditions (e.g. high temperatures and limited substrates), has attracted considerable attention in recent decades. Even under such conditions, water and low molecular weight organic acids (LMWOA) are present. LMWOA play an important role in several biotic and abiotic processes, e.g. in the Krebs cycle used by all aerobic organisms, or as precursors of natural gas in petroleum reservoirs. Further, they can be abiotically and biotically produced. In sedimentary systems, an obvious source of substrates is the water-soluble LWMOA that can be released from organic-rich sediments - however, the controlling factors and rates of generation are still unclear.

So far, most of the studies have focused on short-chain (C1- C4) carboxylic acids, the study of LMWOA being limited due to challenges associated with high volatility. Most applied procedures use derivatization to address this challenge, i.e. conversion of the organic acids into methyl esters. However, this may affect some sample properties, e.g. isotopic footprint. In addition, such procedures also involve the use of hazardous reagents and are time-consuming.

The present study aimed for the characterization of dissolved organic carbon, and its fractions; and, qualification and quantification of underivatized water-soluble LMWOA that can be released from organic-rich sediments as substrates for the deep biosphere. Liquid chromatography - organic carbon detection (LC-OCD) was employed to characterize DOC and its fractions. The methodological strategy unites diverse techniques to encompass the analysis of a wider range of LMWOA, i.e. containing between 1-9 carbon atoms. Ion chromatography was employed to quantify acids with chain lengths between 1-4 carbon atoms. Solid-phase extraction was used as preparatory procedure. SPE extracts were analyzed by GC-MS, to identify the compounds; and GC-FID to quantify saturated acids, i.e. pentanoic, hexanoic, heptanoic, octanoic and nonanoic acid. The investigation also included the testing of effect of water/sediment ratio and phosphate buffer at different pH in the release of

the LMWOA. The final method developed during this study was then successfully applied to 15 samples (shales and coals and oil sand-related sediments) from a variety of depositional environments, with a wide range of total organic carbon (TOC) (2-74%), hydrogen index (HI) (8-703 mgHC/g TOC), oxygen index (OI) (6-207 mgCO2/g TOC), and kerogen type.

It was observed that formic, acetic and oxalic acid are the major components of the aqueous extracts; however, the extracts of shales and coals exhibited differences in acid content. Generally, acid content was higher in the extracts of coals that in those of shales. Pearson's correlation coefficient showed that this result cannot be explained only by the TOC content (higher in coal samples), but also depends on the oxygen index and maturity. Hierarchical cluster analysis was applied to the concentrations of LMWOA (µg/g-sed) in the extracts of shales and coals, and this method identified two groups, which coincided to a classification based on OI.

The wide range of analyzed LMWOA allowed the identification of a pattern distribution. In the extracts of coals, the amount of saturated acids decreased as follows: pentanoic > hexanoic > heptanoic > octanoic > nonanoic acid. However, the shales did not exhibit such a distribution. This outcome is presumed to be related to the parental organic matter and the depositional environment of the sediments.

In addition, this study also showed that the water/sediment ratio used in the method has an effect on the amount of released organic acids. Further, this effect varies depending on the sediment and the properties of organic acid itself. This was demonstrated in the case of oxalic acid in the aqueous extract of a coal sample, whose concentration displayed a dramatic increase when a higher water sediment ratio was applied. pH and phosphate content can also affect the release of LMWOA. When the extraction was undertaken in solutions with pH between 2 to 7, it resulted in an increase in the concentrations of LMWOA. This is probably related to how pH and phosphate affect the charge of the compounds reducing the attractive forces between ions in the solution, favoring the extraction of the acids.

Kurzfassung

Die Forschung zur tiefen Biosphäre, insbesondere das Leben unter extremen Bedingungen (z.B. hohe Temperaturen und begrenzte Verfügbarkeit von Substraten), hat in den letzten Jahrzehnten erhebliche Aufmerksamkeit gewonnen. Selbst unter diesen extremen Bedingungen sind Wasser und niedermolekulare organische Säuren (low molecular weight organic acids – LMWOA) vorhanden. LMWOA sind von Bedeutung in einer Reihe von abischen und biotischen Prozessen – z.B. beim Krebs-Zyklus, der von allen aeroben Organismen durchgeführt wird, oder als Vorstufe von Erdgas in Lagerstätten. Zudem können sie sowohl abiotisch als auch biotisch erzeugt werden. In sedimentären Systemen sind die wasserlöslichen LMWOA eine bekanntes Substrat für mikrobielle Prozesse und können von den an organischem Material angereicherten Sedimenten freigesetzt werden. Die Einflussfaktoren und Freisetzungsraten sind jedoch weitgehend unbekannt.

Bislang hat sich ein Großteil der Forschung auf die kurzkettigen (C1 – C4) Carbonsäuren konzentriert oder auf längerkettige Fettsäuren, die in vielen Studien nach ihrer Derivatisierung zu Fettsäuremethylestern detektiert wurden. Dieses Vorgehen beeinflusst jedoch die Eigenschaften des Analyten, wie beispielsweise das Isotopenverhältnis. Zusätzlich erfordert die Derivatisierung den Einsatz von gesundheits- und umweltschädlichen Reagenzien und ist zeitaufwändig.

Das Ziel der vorliegenden Studie war es, wässrige Extrakte von Kohlen und Schiefergestein zu untersuchen um die Zusammensetzung und die Menge an Substraten für die Tiefe Biosphäre zu charakterisieren. Hier wurden einerseits die wasserlöslichen LMWOA in ihrer originalen, nicht derivatisierten Struktur, qualitativ und quantitativ erfasst, aber auch der gelöste organische Kohlenstoff insgesamt quantifiziert. Die methodische Vorgehensweise kombiniert verschiedene Techniken um ein möglichst großes Spektrum von LMWOA zu erfassen, mit einem Schwerpunkt auf Verbindungen die zwischen ein und neun Kohlenstoffatome enthalten. Für die Quantifizierung der LMWOA mit ein bis vier Kohlenstoffatomen wurde die

Ionenchromatographie eingesetzt. Für die Untersuchung der größeren LMWOA (zwischen C5 und C9) wurde eine Methode entwickelt mit der per Festphasenextraktion die organischen Säuren aus der wässrigen Phase im organischen Lösungsmittel angereichert werden. Die Identifizierung dieser Säuren erfolgte mittels GC-MS, die Quantifizierung der gesättigten Säuren wurde per GC-FID durchgeführt. Mit Hilfe der Molekülgrößenausschlusschromatographie in Kombination mit Kohlenstoffdetektion wurde der gesamte gelöste organische Kohlenstoff in verschiedene Fraktionen unterteilt und quantifiziert. Diese analytischen Methoden wurden auf 15 Proben angewandt. Hier handelt es sich um wässrige Extrakte von verschiedenen Kohlen, Schiefergesteinen, Ölsand und weiteren Proben aus dem Ölsandabbau. Diese Proben haben stark unterschiedliche Gehalte an organischem Kohlenstoff (TOC zwischen 2 und 74%) und repräsentieren verschiedene Ablagerungsmilieus, Kerogentypen und Qualitäten des organischen Materials (HI zwischen 8 und 703 mg HC/g TOC; OI zwischen 6 und 207 mg CO2/g TOC).

Es wurde festgestellt dass Methan-, Ethan- und Oxalsäure die Hauptkomponenten der wässrigen Extrakte darstellen, wobei sich die Schiefer- und Kohleextrakte in ihren Säuregehalten unterscheiden. Grundsätzlich waren die Säurekonzentrationen in den Kohleextrakten höher als in den Schieferextrakten. Der Pearson-Korrelationskoeffizient zeigte, dass dieses Ergebnis nicht allein durch die TOC-Gehalte erklärt werden kann, sondern dass auch OI und Reife des organischen Materials entscheidende Einflussfaktoren darstellen. Diese Korrelation zwischen OI und Konzentration an LMWOA wurde mittels hierarchischer Clusteranalyse aufgedeckt.

Anhand des breiten Spektrums an nachgewiesenen LMWOA konnte ein Verteilungsmuster identifiziert werden. In den Kohleextrakten nahm der Anteil an gesättigten Säuren (C5 bis C9) mit zunehmender Anzahl an Kohlenstoffatomen ab. Für die Extrakte der Schieferproben konnte dieses Verteilungsmuster nicht nachgewiesen werden, was mit der unterschiedlichen Zusammensetzung des organischen Materials erklärt werden kann.

Weiterhin wurde festgestellt dass das Wasser-Sediment-Verhältnis einen Effekt auf die Menge an freigesetzten organischen Säuren hat. Dieser Effekt variiert auch in Abhängigkeit vom Sediment und den Eigenschaften der jeweiligen organischen Säure. Dies wurde für Oxalsäure in wässrigen Kohleextrakten gezeigt, deren Konzentration mit höherem Wasser-Sediment-Verhältnis einen erheblichen Anstieg verzeichnete. Der pH-Wert und der Phosphatgehalt beeinflussen ebenfalls die Freisetzung von LMWOA. Wenn die Extraktion in Lösungen mit pH-Werten zwischen 2 und 7 durchgeführt wurde kam es zu einem Anstieg der Säurekonzentration. Hier zeigt sich die Wirkung von pH und Phosphatgehalt auf die Ladung der Komponenten, die die Freisetzung verstärkt.

Abbreviations

CA	Carboxylic Acids
CDOC	Hydrophilic organic carbon
CoRe	Chemically most complex and recalcitrant fraction of DOC (~1000 Da)
DOC	Dissolved organic carbon
FID	Flame Ionization Detector
GC	Gas chromatography
HI	Hydrogen Index
HMW	Heaviest fraction of DOC (heavier than 20000 Da)
HOC	Hydrophobic Organic Carbon
IC	Ion chromatography
INT	Intermediate fraction of DOC (between 300 and 500 Da)
IS	Internal Standard
LC-OCD	Liquid chromatography-organic carbon detection
LLE	Liquid-liquid Extraction
LMW-acid	Acidic fraction of DOC (lower than 350 Da)
LMW-neutral	Neutral fraction of DOC (lower than 350 Da)
LMWOA	Low Molecular Weight Organic Acids
MFT	Mature fine tailings
MS	mass spectroscopy
OI	Oxygen Index
OS	Oil sand
r	Pearson correlation coefficient
R^2	Coefficient of determination
R_0	Vitrinite Reflectance
SPE	Solid-phase extraction
Tmax	Temperature of maximum Pyrolysis Yield
TOC	Total Organic Carbon
TS	Tailings sand

Contents

4. Methodological development

5. Characterization of DOC and LMWOA in aqueous extracts from organic-rich sediments

6. Characterization of LMWOA that are released from organic-rich sediments

Index of figures

Index of tables

1. Introduction

For a long time, scientists believed that extreme conditions such as high pressure, lack of oxygen, and low supply of nutrients and energy would make deep, subsurface environments uninhabitable for any life form. Nonetheless, investigations have proven the existence of the deep biosphere (Gold, 1992; Pedersen, 1993).

The study of the deep biosphere includes the identification and the quantification of microbial communities, and the metabolic pathways that allow the survival. A decrease in microbial abundance with depth, and a strong correlation to mean sedimentation rate and distance from land have been reported. Such an abundance varies between sites by ca. five orders of magnitude (Kallmeyer *et al*. (2012). Bacteria and archaea have commonly been found in the subsurface. Sulfate-reducing bacteria (SRB), inhabitant of hydrothermal vent sites, utilize sulfate as the terminal electron acceptor but many can also use other sulfur compounds, nitrate, and fumarate. As electron donors, SRB oxidize different compounds, mostly using low-molecular weight organic species such as lactate, pyruvate, ethanol, short-chain carboxylic acids or hydrogen (Dhillon *et al*., 2003).

Alongside the questions about the microbial communities and metabolic pathways, the carbon sources available as feedstock remain unclear. As water is ubiquitous, the water-soluble compounds that can be released from organic-rich sediments at reservoir temperatures, may supply the necessary substrates for the deep biosphere. In natural waters, the sources of the organic matter has been studied by means of the analysis of dissolved organic carbon (DOC) (Thurman, 1985; Thomas, 1997; Sachse *et al*, 2001; Imai *et al*, 2001; Sachse *et al*., 2005). DOC is defined as the organic carbon smaller than 0.45 μm in diameter; and it includes free amino acids (FAA), peptides, proteins, polysaccharides (PS), carboxylic acids (CA), nucleic and humic substances (HS). DOC has also been divided into categories based on polarity, i.e. hydrophobic, hydrophilic, acidic, basic, and neutral (Sachse *et al*, 2001; Imai *et al*, 2001; Sachse *et al*, 2005). Similarly to natural waters, the study of DOC and its fractions in the water extracts of shales and coals, may provide at insight into the relationship between organic matter and other properties. Furthermore, short-

chained carboxylic acids are well-known substrates for microbes and have been reported to be abiotically or biotically produced. Thus a characterization of the water-soluble organic acids might provide a deeper understanding of the processes related to the potential carbon and energy sources.

1.1 Presence and roles of LMWOA in geology

Carboxylic acids (CA) are compounds characterized by the presence of at least one carboxyl group, and exhibit different structure and bonds, that goes from saturated, monoenoic (trans acids), methylene-interrupted polyunsaturated, bis- and polymethylene-interrupted, conjugated, up to branch-chain and cyclic acids, and may contain more functional groups as well. The properties of CA depend on this diversity, which in turn impact their functions in environment. For example, in living cells, CA are important since they are related to metabolic pathways, they participate as intermediates in the tricarboxylic (TCA) cycle in aerobic organisms, or carrying out storage and membrane functions in form of lipids. Although over 1000 CA are known, only around 20 occur widely in nature, most having even chain lengths between C4 to C22, C18 being the most common (Gunstone, Harwood, & Dijkstra, 2007). Low molecular weight organic acids (LMWOA), having chains with one to 12 carbon atoms, are practically ubiquitous in nature and take part in several biotic and abiotic mechanisms as well.

The presence and roles of LMWOA in geology were mainly result of the research on fossil fuels. In order to examine the origin of natural gas, Carothers and Kharaka, (1978) carried out a characterization of short-chain carboxylic acid anions in 95 water samples from oil and gas fields in California and Texas. They reported that concentrations of acetate, propionate and valerate depended on the temperature of formation waters. Briefly, in zone I temperature below 80 °C, the concentration of CA was up to 60 mg/l and propionate was predominant; zone II was characterized by temperatures between 80 up to 200 °C and the acid content was up to 4900 mg/l. Zone III practically lacked CA. Based on this study, the authors suggested that origin of carboxylic acid anions in formation waters could be due to thermal decomposition of kerogen, which may in turn endure decarboxylation in reservoir rocks producing

natural gas. Fisher, (1987) detailed the temperature range reported by Carothers and Kharaka (1978). He reported a study of aliphatic acid anions in 144 formation waters from 8 places covering a temperature range between 40 up to 100 °C. The author concluded that zone I extends from 20 up to 90 °C and the amount of CA increases as temperature raises, with acetate ion dominating. In zone II, the acid anion concentrations reduces in relation to temperature, in which it could be between 90 and ca. 250 °C, supporting the idea that decarboxylation may be involved. In general acetate and propionate were more abundant than butyrate, whose concentration in turn was greater than that of valerate. Another contribution of this work was that concentrations of carboxylic acid anions are not only temperature-dependent, but water mixing also plays a role.

The understanding of secondary porosity in reservoirs is important to study fluid pathways. In their work on the chemistry of second porosity (porosity enhancement), Surdam *et al.,* (1984) carried out experiments of kerogen maturation in aqueous environment for type I and III kerogens. They demonstrated that some carboxylic acids (acetic and oxalic acid) are produced from both kind of kerogens (algal and humic materials). The process also seemed to be pH dependent. In similar environmental conditions as reported by Carothers and Kharaka (1978), Surdam *et al*. concluded that oxalic acid has a larger effect on enhancement of porosity (by an increase of dissolution of aluminium) than acetic acid. Moreover, the organic matter content (thus LMWOA), composition and the magnitude of diagenetic reactions will also determine the extent of such an enhancement.

As hydrous pyrolysis was a common procedure to examine the hydrocarbon production in source rocks and then extended to the research of carboxylic acids, Lundegard and Senftle, (1987) evaluated the suitability of this technique in the release/ study of organic acids (C1 – C4) from several kerogens. During heating tests, they reported that acetic and propanoic acid were major products, with little amounts of formic acid. The sample with lowest TOC yielded the greatest amount of C1 - C4 acids, normalized to C content in the unreacted kerogen. In case of acetic acid, its generation was positively correlated to temperature, but it was highest in the rock with the lowest TOC, which coincidently exhibited the highest OI. Their

experiments also showed that acetic acid production was still higher even if the decrease of propionic acid was insignificant, which proved that acetic acid formation is unrelated to propionic acid destruction, as suggested previously. Tests assessing reaction time showed that generation of acetic acid increases as time goes by until it stabilizes. However, the rate of decarboxylation of propionic acid overcame its synthesis. In conclusion, results indicate that conditions under which hydrous pyrolysis is usually conducted, lead to loss of some acid species, thus an underestimation of the amount of those acids may have occurred.

Once the role of LMWOA in geology was recognized, questions arose about their origin. Kawamura *et al.,* (1986) reported the generation of carboxylic acids from Type I and II kerogens derived from pyrolysis and hydropyrolysis and the role of montmorillonite, illite and calcite. The authors identified aliphatic and branched acids and benzoic acid with chain length between 1 and 10 carbons. In pure- kerogen extracts, the major pyrolysis product was acetic acid followed by formic and propionic acid, with the remaining acids (C4-C10) exhibiting a slightly even/odd carbon preference. These authors suggested that CA may come from desorption (CA had previously been adsorbed and trapped in the kerogen); from unsaturated fatty acids and their degradation and/or polymerization products, or deamination of amino acids and peptides, all of these chemical reactions being catalyzed by high temperatures. Helgeson *et al.*, (1993) investigated the hydrolysis of hydrocarbons at the oil-water interface as origin of LMWOA found in oil field waters. The authors calculated the distribution of species with minimum Gibbs free energy in oil field waters in the Texas Gulf Coast under the assumption of a metastable equilibrium between representatives of organic and inorganic aqueous species at reservoir temperatures and pressures, calcite and albite. Briefly, they used the log f_{O2}[1] as thermodynamic parameter to characterize the oxidation/reduction reactions amongst both inorganic and organic species. Based on these calculations, Helgeson *et al.* (1993) proposed that water may directly react with alkyl carbons. They claimed that alkanes with a maximum of 9 carbons in the chain length can react with water under reservoir conditions in the subsurface (i.e., ~100 to 150°C and 40 MPa) to form CO_2 and

[1] Fugacity of oxygen (f_{O2}). Albarède (2011) defines oxygen fugacity as an equivalent of the partial pressure of oxygen in a particular environment (atmosphere, rocks, etc.) corrected for the nonideal character of the gas.

alkanes with one less alkyl group[2]. This would could produce an oxygen excess and an external source of hydrogen from water. In a subsequent reaction, water could react with free-radical sites on alkyl groups to form an aldehyde. In presence of water, the aldehyde in turn could transform into a carboxylic acid or CO_2.

Dias *et al.*, (2002) studied the isotopic footprint of LMWOA during pyrolytic process on shales, limestones and a lignite, to determine the origin of the acid content. The authors performed hydrous pyrolysis under a wide range of temperatures on six immature source rocks of type I, II, IIS, and III kerogen to identify the isotopic footprint of possible mechanisms that generates CA. They reported the predominance of acetic acid over other acids irrespective of the differences in kerogen type and mineral content. The authors also observed that ^{13}C content of carboxylic acids inversely relates to molecular weight, but positively to thermal maturity. Dias *et al.,* (2002) concluded that generation mechanisms such as pyrolysis of carboxylic acids, oxidation of suitable acid precursors (e.g. esters), the release of trapped C, ionic-bond break down or decarboxylation cannot singly explain the observed isotopic variations. They suggested that once the organic acid are generated, additional reactions with dissolved inorganic C are involved.

Seewald (2001) provided an approach of how LMWOA are abiotically produced in basinal brines. This author proposed that LMWOA derive from aqueous aliphatic hydrocarbons following an oxidative pathway. *n*-alkanes are oxidized to the homologue alkenes; then, with addition of water, *n*-alkanols are formed, which undergo oxidation to form *n*-alkenones. In presence of water, the C=C bond of the *n*-alkenone is broken resulting in two carboxylic acid molecules of decreased chain length. This model also explain the predominance of acetic acid over other acids, namely, acetic acid is produced during the degradation of several hydrocarbons with different chain length. The model even applies to short-chain hydrocarbons, however, it is restricted by stoichiometric limitations.

Another study was conducted by Bertilson & Tranvik (1998), who demonstrated that LMWOA can be photochemically produced. In opposite to previous works, they used natural water samples from humic lakes (subject only to filtration) to assess the effect

[2] Alkyl refers to an alkane missing one hydrogen, e.g. methyl (CH_3-) , acetyl (CH_3CH_2-).

of irradiation (artificial, though comparable to natural intensities, and natural sunlight) in the production of formic, acetic, oxalic and malonic acid. The water sample with higher dissolved organic carbon (DOC) exhibited an important production of carboxylic acids, formic acid being the major degradation product. They also determined that the rates of production of LMWOA were similar to photochemical production of CO_2 and algal primary production, thus it can be recognized as a relevant. In addition, they concluded that the microbial oxidation of carboxylic acids appears to be an important source of energy, though they are poor in energy compared to sugars for example.

Summarizing, up to 22 CA have been identified in pyrolysates of kerogen; however, the application of hydrous pyrolysis could have led to an underestimate of the concentrations of some compounds (e.g. propanoic acid) released from kerogen. The high temperatures used in pyrolysis probably cause thermal decomposition of some carboxylic acids. Furthermore, LMWOA can be either biotically or abiotically produced. The biological production may involve degradation of hydrocarbons either under oxic or anoxic conditions. Abiotic processes include photoxidation of humic substances in the water column and the oxidative degradation of aliphatic hydrocarbons. However, isotopic analyses have shown that mechanisms behind of the formation of LMWOA can relate to complex reactions also involving the role of inorganic carbon.

LMWOA and the deep biosphere

The study of LMWOA in the deep biosphere started not long ago and has implications for many other areas of research e.g. methane production (Stams, 1994), biochemistry of aquifers in granitic rocks (Pedersen, 1997), calcium carbonate precipitation (Bosak & Newman, 2005), and, degradation of petroleum in reservoirs (Head *et al.*, 2003) and its resulting economic impact.

McMahon and Chapelle (1991) studied the concentration of formate and acetate in pore water of aquitards (with little respiratory activity) and aquifers (actively sulphate-

reducing). These authors showed that acid content was greater in the former (55 – 110 µM) than in the latter (30 – 35 µM), likely due to microbial activity. Such a difference would result in diffusive processes of acids from aquitards (high concentration) to aquifers (low concentration). They also carried out incubation experiments comparing the acid content in pore water of both types of sediments and found that in aquifers concentration of both acids stayed at low levels, whereas it was higher in aquitards, with no acid accumulation in the pore water. This suggested that organic acids were being produced, but consumed at a very low rate. The authors proposed that aquifer sediments exhibited anaerobic respiration by sulfate reduction whereas the aquitards were fermentative.

Chapelle & Bradley (1996) conducted analysis of acetogenic activity in deep sediments. They examined pore water of sediments (90 – 888 m depth) of three core holes in South Carolina and Georgia (USA). Concentration of acetate was 1800 µM (108 mg/l), whereas concentration of formate was as high as 6400 µM (294 mg/l). The authors reported that concentrations of formate and acetate were different among the three core holes, but generally the acid content was higher in pore water of clays in comparison to those of sands and silty sands, which exhibited lower amounts of acids. Autotrophic acetogenic activity was also proven in the sediment cultures - the microbes isolated from one of the samples were able to grow under three tested environments (autotrophic, heterotrophic and mixotrophic media) resulting in the production of formate and acetate. These microorganisms were identified as gram positive. Based on these results, this study showed that CA could be produced biologically at low temperatures; furthermore, the authors highlighted that in the studied core holes, the generation of CA was due to microbial activity rather than abiotic.

In order to investigate the circulation and fate of LMWOA in shallow contaminated aquifers, Cozzarelli *et al.*, (1994) studied the geochemical evolution of LMWOA, in the direction of groundwater flow (downgradient) from a crude-oil spill near Bemidji, Minnesota. The authors divided the detected LMWOA in aliphatic, alicyclic and aromatic groups. The concentrations of organic acids downgradient from the oil body increased over time. The aliphatic fraction exhibited a remarkable increase; further,

acetic acid increased exponentially. Cozzarelli *et al.*, (1994) concluded that the increase near of oil lens was due to microbial degradation, mainly methanogenesis. Over time, Fe^{2+} available to iron reduction depleted, thus methanogenesis became more important. Based on this and the degradation kinetics of aromatic hydrocarbons, the authors pointed out that acetoclastic methanogenesis may be the rate- limiting step in the anaerobic degradation of the hydrocarbons. As supported by previous works, conversion of acetate to methane may assist the degradation of aromatic compounds by allowing a syntrophic (cross feeding) association between fermentative and proton-reducing microorganisms and methanogens.

LMWOA importance is also related to the exploitation of oil. SRB activity increases sulfide concentrations (souring) in oil reservoirs, which represents a problem during the secondary oil recovery to produce the remaining oil. Injection of nitrate into the reservoir stimulates the growth of nitrate-reducing, sulfide-oxidizing bacteria (NR-SOB) and heterotrophic nitrate-reducing bacteria (hNRB) that compete with SRB for degradable oil constituents (Hubert & Voordouw, 2007). Grigoryan *et al.,* (2008) studied the effect of acetate, propionate, and butyrate as substrates to SRB and hNRB in an oil field in Argentina. The authors found out that SRB first consume propionate and butyrate producing acetate. Acetate is only used when the concentration of this acid reaches a certain value. However, hNRB uses the acetate, propionate, and butyrate in similar rates reducing nitrate to nitrite and nitrogen. When the concentration of nitrite is ca. 1mM, the produced nitrite can inhibit SRB activity, additionally leading to an accumulation of acetate.

Geological disposal is seen as a safe solution for the long-term management of radioactive waste. Crystalline rocks, salt and clays are usually considered for geological disposal. Argillaceous formations are composed of clay minerals, carbonates, quartz, feldspar, pyrite, and small amounts of organic matter (depending on depositional conditions). Chautard, *et al.,* (2015) studied the pore water in a Toarcian Argillite in order to examine the effect of extractant pH, temperature, and oxygen exposure, on the dissolution of LMWOA under anoxic conditions (supposed to occur in the disposal site). The concentrations of formate, acetate, priopionate, lactate and fumarate were reported. The authors concluded that an increase of pH

from 7 to 9, results in an increase in the release of LMWOA. At 25 °C, the LMWOA concentrations were lower when compared to the release at 60 °C. When the sediment was exposed to ambient air for 2 h, the acid content increased. The analysis of LMWOA shows that during the operation of a geological disposal site of radioactive waste, there are conditions (e.g. gallery excavation) that may trigger the bacterial activity, thus affecting the formation of complexes, which may modify the mobility of compounds as rare earth elements (REE), heavy metals and radionuclides in the disposal site. REE are produced during radioactive decay and the knowledge of their behavior may serve as a model for actinides, for which thermodynamic data are difficult to obtain. Heavy metals have been proven to be toxic for human beings, plants, animals and many microorganisms.

Deepening the study of coals, Bou-Raad *et al.*, (2000) assessed the remotion of undesirable inorganic compounds using simple aqueous extraction with a Soxhlet apparatus. Additionally to the inorganic compounds, these authors reported that acetate and oxalate were also extracted. Based on this paper, Vieth *et al.,* (2008) evaluated the effect of thermal maturity on the potential release of LMWOA in New Zealand coals, mundstones and sandstones, as feedstock for the deep biosphere. They reported higher amounts of formate in the water extracts compared to those of acetate and oxalate. The latter were even detected in water extracts of mudstones and sandstones (0.2 – 5.2 % TOC); furthermore, they stated low-rank coals released higher amounts of CA, thus maturity has an negative effect on the potential of a sediment as feedstock. Additionally, based on Horsfield *et al.*, (2006) calculations, they concluded that the amount of released LMWOA should be enough to feed deep biosphere communities for several thousand or even millions of years. Glombitza *et al.*, (2009) determined the amounts of formic, acetic and oxalic acid released by alkaline ester cleavage reaction from coals belonging to the New Zealand Coal Band. This Band comprises a series of coals with a wide and almost continuous maturity (R_0 between 0.25-3.0%) from Cretaceous to Tertiary age. The cleavage was applied to isolated kerogen, and the determination of LMWOA involved the dissolution of potassium salts of acids (cleavage products) in water to be analyzed by IC. The authors estimated generation rates of formic, acetic and oxalic acid to determine if the amount of the released LMWOA was enough to meet the carbon requirements of

the deep biosphere. They concluded that the coals released amounts of ester-bound LMWOA (to the kerogen) in enough quantities to support deep microbial life over geological time.

1.2 Analytical methods for the analysis of LMWOA

Several methods have been developed to identify and quantify carboxylic acids in geological samples. Some illustrative examples used on geological samples (liquid and solid phase) and soils are summarized below to provide a background for this study. These highlight the importance of laboratory procedures in the efficiency of extraction, types of LMWOA extracted, and the simulation of natural processes. It also is useful to point out that underivatized short-chain carboxylic acids, matter of this study, are difficult to analyze in gas chromatography due to their reactivity and volatility, thus it has been a common practice to transform them into methyl esters (Zoccolillo *et al.* 1982). However, this operation relates to isotopic fractionation (Rieley 1994), which represent a disadvantage in the study the origin of the acids. Thus a method capable of determining underivatized LMWOA is desirable.

Carothers and Kharaka (1978) carried out the characterization of short-chain carboxylic acid anions in 95 water samples from oil and gas fields in California and Texas. Amounts of LMWOA were determined by concentrating the salts of acids down to dryness using a rotary evaporator. The solid remains were dissolved in distilled water, which was purified using an ion-exchange resin. GC-FID was utilized to analyze the final aqueous solution. Surdam, Boese, & Crossey, (1984); MacGowan & Surdam (1988); MacGowan & Surdam (1990) employed ion chromatography (IC) to determine carboxylic anions (between formate and octanoate) in the aqueous samples equipped with a conductivity detector.

Barth (1987) developed a different procedure based on capillary isotachophoresis (ITP). Briefly, in ITP the solution is exposed to an electric field which will move the

dissolved charged ions with fixed velocity (isotachophoretic condition). Ion mobility depends on the ion charge, shape and size, the characteristics of the medium and the electric field applied; assuming ions as spheres, ionic mobility will be inversely proportional to the particle size (Karol & Karol, 1978). As ions with equal mobility may overlap, specific media (electrolytes and mixtures) are needed to allow effective separation. For analysis, Barth utilized a tachophore which included a column with two detectors: a conductivity detector and an UV detector at 254 nm. Both signals give form to the tachogram, it shows the analytes on a ladder-type trace (Fig. 1.1). For identification and quantification, this method requires calibration curves of standard solutions containing the target analytes. Slight changes in the electrolytes can affect the position of acids on the tachogram, thus, the use of internal standard addition or co-injection are preferred. The method by Barth consisted in a 30 - min qualitative analysis (including a pre-separation step) under two different pH (4.3 and 5.5) and 5 electrolyte mixtures, and calibration curves of standard solutions with 29 acids (saturated, hydroxy, cyclic and aromatic, mono and dicarboxylic acids). This method was applied to analyze formation waters and produced waters samples from Norwegian Continental Shelf (Barth, 1987); aqueous extracts derived from hydrous pyrolysis of sources rocks and crude oils (Barth *et al.,* 1988; Barth *et al*., 1989; Andresen *et al*., 1994).

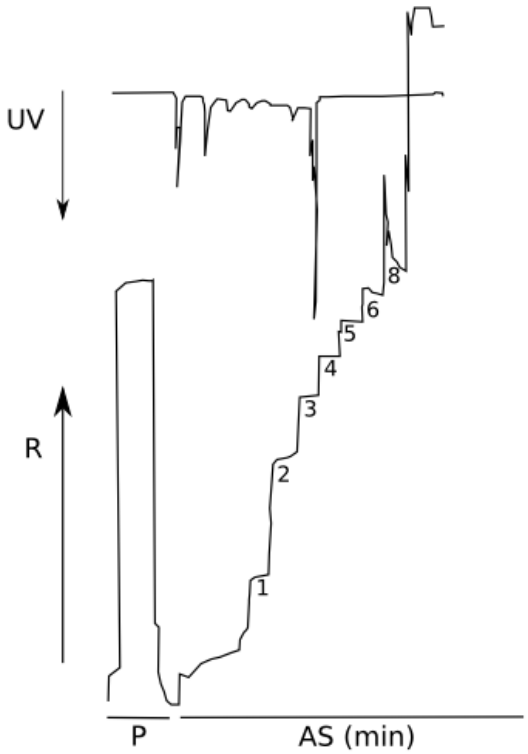


Fig. 1.1 Reproduction of a tachogram published by Barth (1987). It shows the UV and conductivity traces of a standard solution of saturated monocarboxylic acids. See the text for explanation. Nomenclature: 1 – 8, carbon number chain length of acid; P, pre-separation period (in min); AS, analytical separation in minutes; R, zone resistance (conductivity detector); UV, absorbance at 254 nm.

In order to increase the number of low molecular weight aliphatic carboxylic acids (LACAs) that could be determined in paddy soils and to provide a more complete overview of the amounts of 10 LACAs in several soil types in Japan, Tani *et al.,* (1993) developed a method. This consisted of three steps: a) extraction with water or ammonium phosphate buffer by means of a reciprocating shaker and ultrafiltration of the extract (0.025 µm); b) purification and concentration. Phenolic compounds were removed using a polyvinyl pyrrolidone resin in batch mode; subsequently, continuous liquid-liquid extraction with diethyl ether was applied to the extract which was concentrated by evaporation. This concentrate was loaded on to a cation-exchange column; c) quantification was done by HPLC analysis. Acids were identified by means of standards which have undergone the same procedure as the samples. The extraction using two different solvents (water or ammonium phosphate buffer) provided a comparison between water-soluble and absorbed CA in the soils. The amounts of LACAs obtained via phosphate buffer were higher than those water-

soluble, and five types of CA represented near 90 % of the amounts: formic, acetic, oxalic, succinic and citric acid. The authors pointed out that the acetic was the major component in the extract, demonstrating the advantage of this method to detect this acid in comparison to previously used methods. This method was applied to determine the amounts, compositions, and distribution and season changes of LACAs in several forest soils in Japan, e.g. andisols, acid brown forest soils, and podzols. However, this procedure involved a complex home-made apparatus, long running time for HPLC analyses and subsequent clean-up. Thus, Tani and his group developed a new method using IC. The modified method by Tani *et al.*, (2001) consisted in soil samples and ultra-pure water placed in polypropylene vessel and shaken for 1 h. The solution was centrifuged and filtered. Subsequently, the aqueous extract was concentrated by freeze-drying and supernatant was a alkalized with NaOH. The residue was again dissolved in ultra-pure water by ultrasonification. Cation content and colored hydrophobic organic matter are removed from the concentrate by using a cation-exchange cartridge (Sep-Pak Pluss Accel CM) and a reverse-phase cartridge. (Sep-Pak Plus C8 or Sep-Pak Plus tC18) respectively. The LMWOA are then identified and quantified by IC on a TSK gel Oapak-A column and a guard column, TSKgel OApak-P. A TSKgel IC-ANION-Pwxl column was used to determine oxalic acid. Formic, acetic, propionic, butyric, lactic, oxalic, malic, and citric acid were detected using this method.

Farrington and Quinn (1971) compared several procedures to extract fatty acids in young sediments and concluded that saponification was the most effective technique to extract FA amongst the following: soxhlet extraction, sonication, direct solvent reflux, and the one developed by Caret and Dyer (1959). Based on the results of these authors, Kawamura & Ishiwatari (1981) employed a strategy involving a pre-extraction step (with a benzene/methanol mixture), saponification of the organic extract (to quantify unbound FA) or extracted sediment (to determine bound FA) with KOH-MeOH and purification of the supernatant with hexane/ether to remove neutral compounds. Then, the extract was acidified and a second liquid-liquid extraction (LLE) was applied to recover the FA using a hexane/ether mixture, followed by a methylation step with BF3/methanol. Prior to GC-FID analysis, the methylated FA were purified by column chromatography. They observed that heating favored the

extraction of unbound of FA with carbon chain length of 12 to 19, and both types (bound and unbound FA) from C20 up to C30, had also increased with the rise in temperature.

Eglinton *et al.,* (1987) also carried out hydrous pyrolysis experiments on immature kerogens using a different methodological approach to analyze a wider range of acids. The pyrolysis solution was purified by means of LLE and DCM, MeOH and light petroleum as solvents. The remaining aqueous extract was centrifuged and the supernatant was filtered, and $HgCl_2$ was added in order to prevent microbial degradation. Subsequently, the pH of aqueous solution was raised up to 8-9 and concentrated by evaporation, the residues were dissolved into BF_3/*n*-butanol, and esterification was achieved by heating the mixture in a water bath. Another cleaning step was carried out to finally dissolve the ester in DCM to be analyzed by GC and GC-MS. With this method, 22 different acids (lineal, branched and aromatic acids, mono- and dicarboxylic acids) were identified with chain length between C2 and C10.

Solid-phase extraction (SPE) and headspace solid-phase microextraction (SPME) coupled to gas chromatography – mass spectrometry (GC-MS) or gas-chromatography with flame ionization detector (GC-FID), have been also employed to determine concentrations of carboxylic acids in water samples. Ábalos *et al.* (2000) optimized a SPME method to determine underivatized saturated carboxylic acids, with carbon chain length between 2 and 7, in waste water samples. In SPME, a fused silica optical fiber coated with a thin layer of a polymer extracts analytes from a liquid, the headspace above a liquid or solid, or a gaseous phase. The fiber subsequently is introduced into a gas chromatograph (GC) and the analytes are thermally desorbed into the hot injector. The procedure developed by the authors employed a polydimethylsiloxane–Carboxen (PDMS-CAR) fiber with 20-min extraction time at 25 °C, and vigorous agitation of the aqueous phase. The desorption in the GC injector was 5 min at 300 °C. Examining the isotopic footprint of CA during pyrolytic process, Dias *et al.*, (2002) applied acid-base liquid-liquid extraction to remove non-polar compounds in the aqueous pyrolysate, followed by an extraction of CA using solid-phase micro extraction (SPME). GC-MS on a FFAP column was employed for identification and quantification of the acids with carbon

length chain between 2 and 6. Using this approach, the authors concluded that generation of organic acid may involve a more complex series of mechanisms (e.g. reactions with dissolved inorganic C). Pyrolysis of carboxylic acids, oxidation of suitable acid precursors (e.g. esters), the release of trapped C, ionic-bond break down or decarboxylation cannot singly explain the observed isotopic variations.

In SPE, a stationary phase (sorbent or resin) extracts or adsorbs one or more components from a liquid phase (sample). Subsequently, the analytes are recovered by washing the phase with a solvent, known as eluent, which in turn is analyzed by a GC. Jurado-Sánchez *et al.*, (2010) tested the adsorption of 22 carboxylic acids onto 14 different SPE phases for concentration and separation of carboxylic acids from an aqueous matrix. The atom number of the carbon chain length ranged between 2 to 18, the list included linear, branched, unsaturated, aromatic and dicarboxylic acids. Also three GC detection methods were assessed for quantification, namely flame ionization detection (FID), mass spectrometry with electron impact (EI) and positive chemical ionization (PCI). Because two SPE phases showed the best sorption efficiency, the combination of both products was evaluated to increase the recovery. The procedure described as being the most successful involved the simultaneous use of two reversed solid phases Supelclean ENVI-18 and LiChrolut® EN in a ratio of 1:1, methanol as eluent and 2-*tert*-butyl-4-methylphenol (2tb4mP) as an internal standard. Supelclean ENVI-18 exhibited an efficiency higher than 85 % for medium and long chain aliphatic and aromatic acids, whereas LiChrolut EN performance was best for acids with carbon atom number from 2 up to 12 and aromatic acids. Briefly, the procedure began with the acidification of the sample, which was loaded through a home-made SPE column with a flow rate of 4 ml/min provided by a peristaltic pump. After the sorption of the acids, the column was dried with air stream for 2 min with an inverted flow direction. Subsequently, 200 µl of methanol spiked with the internal standard was injected to recover the acids (elution step). The organic extracts were analyzed by GC-MS equipped with an INNOWax column using external calibration for quantification. This indicates that aqueous standard solutions should be used to build calibration curves for each acid and the concentration of individual compounds in extracts of natural samples should be calculated by interpolation.

1.3 Motivation and approach

In geochemistry, LMWOA have been investigated from different perspectives based on the objectives of individual studies. Their roles include natural gas precursors in the petroleum system; as elements in reaction pathways in clastic sediments, which determine the development of a secondary porosity, relevant to petroleum migration; and, as substrates and degradation products for microbes under aerobic and anaerobic environments.

As described lines above, most studies have focused on saturated acids with carbon chain length between 1 and 4, and 14 and longer chains; however, there is very little information in the amounts and sources of pentanoic, hexanoic, heptanoic, octanoic and nonanoic acid. Thus, the aim of this thesis was to study the potential of geological samples to release organic compounds, focused on determining the concentration of carboxylic acids and the impact of TOC, maturity, kerogen type, and depositional environment on their concentrations. Further, this study also sought for a method that does not require conversion of acids in FAME and capable of working with complex samples with limited laboratory preparation. In analyzing the samples using the currently most popular and established analytical methods, a variety of new challenges (e.g. low recovery and degradation of organic extract) were identified. As this optimized method is applicable to samples from a variety of geological environments, it has the potential to be used as a standard method.

The present project, “The characterization of water-soluble organic compounds released from organic-rich sediments”, has the following objectives:
- identifying the type of compounds that are released from organic-rich sediments when subjected to subsurface conditions, i.e. heating in presence of water, similar to those prevalent in deep biosphere. This also includes the determination of the DOC and its fractions in the aqueous extracts;
- determining the amounts of underivatized carboxylic acids with carbon chain length from 1 up to 9 that can be released from shales and coals;
- evaluation of the controlling factors (e.g. OI, kerogen type) on the extracted quality

and quantity of LMWOA

- testing conditions that may impact the release of the acids in selected samples.

In order to carry out such a characterization an analytical strategy was designed, which required the analysis of underivatized analytes. The analysis of underivatized carboxylic acids, which circumvented the possibility of isotopic fraction, was crucial for this undertaking. The analytical strategy involved several parts:

1) An experimental model. This simulates the environment in which LMWOA are released, specifically presence of water and temperature (80-90 °C);

2) Sample preparation. This step allows adapting the sample to enable processing by analytical techniques (e.g. GC-MS). This may include chemical and physical processes as concentration and acidification.

3) Analytical techniques. In this step the prepared sample is measured to identify and quantify the target analytes. The applied techniques were chosen regarding the nature of the acids, C1-C4 acid were determined by IC, and SPE-GC-FID/MS was employed to characterize water-soluble compounds with longer carbon chain lengths.

4) Processing data. It comprises calibration curves, chromatogram interpretation and statistical evaluation.

This thesis is divided in several chapters.

Chapter 2 (Samples) presents a description of the shales and coals chosen for this project. The sample set consists of immature shales and coals with a TOC content between 2% and 74%. The samples represent four different kerogen types, namely I, II, III and II/III.

Chapter 3 (Methodology) describes the procedures applied and the operational conditions under the measurements were performed.

Chapter 4 (Methodological development*)* explains the methodological approach and the modifications that should be carried out due to a series of difficulties. For example, the unsuitability of one-point calibration for the quantification of underivatized LMWOA by GC-MS and GC-FID; and the degradation of the SPE extract, which represented a major hindrance.

Chapter 5 (Characterization of DOC in aqueous extracts from organic-rich sediments) describes DOC content and its fractions, in the aqueous extracts.

Relationships between DOC, TOC content, geological age and maturity are also discussed.

Chapter 6 (Characterization of LMWOA in aqueous extracts from organic-rich sediments) describes the types of carboxylic acids and other water-soluble compounds identified by IC and GC-MS; and, the quantification of formic, acetic, propionic, butanoic, pentanoic, hexanoic, heptanoic, octanoic, nonanoic, oxalic, pyruvic, malic and methylsuccinic acid, by IC or SPE-GCFID. Relationships between TOC content, acid content, geological age and maturity are also discussed. It also details parameter-based tests that show how the release of the LMWOA can potentially be affected by the water-sediment ratio and pH-complexation. In this chapter, the potential of the studied organic sediments as carbon sources for the deep biosphere is also examined.

Chapter 7. In this, the findings of the present work are summarized and conclusions are drawn.

The Appendices are located at the end of the thesis.

2. Samples

Geological samples representing a variety of depositional environments, with a wide range of total organic carbon (TOC) (2-74%), hydrogen index (HI) (8-703 mgHC/g TOC), oxygen index (OI) (6-207 mgCO2/g TOC), and kerogen type were selected. The samples include:

- **Shales** include a sample from Botneheia formation, with HI of 249 mHC/g TOC and kerogen type II/III; another from Schöneck fm with a HI of 540 mHC/g rock; a Posidonia shale from Wickensen well with a high Hydrogen Index (HI) (635 mHC/g TOC) and OI of 20 mgCO2/g TOC; a black shale from Irati formation representing kerogen type I and the highest HI and lowest OI among the samples, 703 mHC/g TOC and 6 mCO2/g TOC, respectively; a mature Alum shale completes the set of shales.
- **Coal**: A Cannel Coal (kerogen type I/III), five coals from New Zealand (kerogen type III), and a German coal from Lausitz (kerogen type III) conform the set of coal. In this set, OI varies between 6 to 77 mgCO2/g TOC and the HI ranges from 109 to 211 mHC/g TOC. Besides coal and shales,
- **Oil-sand related samples**: the sample set also includes an oil sand (OS), a mature fine tailing (MFT), and a tailings sand from Alberta, Canada, which TOC content ranges from 3.75 to 15.4 %.

Table 2.1 and Fig 2.1 show basic characteristics of the selected samples and the available data of Rock-Eval pyrolysis. For some samples characterization of kerogen via petroleum type organofacies (Horsfield, 1989; Horsfield, 1997) (Fig. 2.2) and aliphatic, aromatic, oxygen- and sulphur-bearing compounds (Larter, 1984) was available. A detailed description of these samples is presented below.

Botneheia shale (G001799)

This sample is one of the Norwegian Geochemical Standards (NGS), Svalbard Rock-1 (NGS SR-1) with a relatively low TOC content (2.17%), a vitrinite reflectance (Ro) of 0.41 (%), is kerogen type I/III as suggested by the Rock-Eval pyrolysis data (Fig. 2.1). Svalbard is well known for its geological diversity, with sections representing most of the Earth's history, and depositional environments ranging from highly anoxic, carbonate-rich shales to shallow marine anoxic toward south. Mørk and Bjorøy (1984) reported that Mesozoic sediments in Svalbard had reached a suitable maturity for hydrocarbon production, especially those belonging to Lower to Middle Triassic Sassendalen Gp, which consists of marginal to deeper organic-rich marine clastic sediments.

Botneheia Fm is part of the Sassendale Gp which comprises Lower and Middle Triassic sediments rich in pelagic fossils (e.g. ammonoids) and phosphate nodules (Dallmann, 1999; Krajewski, 2008), that can be classified as shales, silstones and sandstones. It records a coarsening-upward succession (Dallmann, 1999) and the sediments are characterized by organic rich shales, up to 10 % of TOC, which were deposited in a deep shelf environment between the Olenekian to the Ladinian (Riis *et al.*, 2008). The rich organic carbon content of the sediments is the result of high organic productivity during deposition (Krajewski, 2008), and is also related to anoxic/dysaerobic conditions (Mørk and Bjorøy, 1984; Mørk and Worsley, 2006).

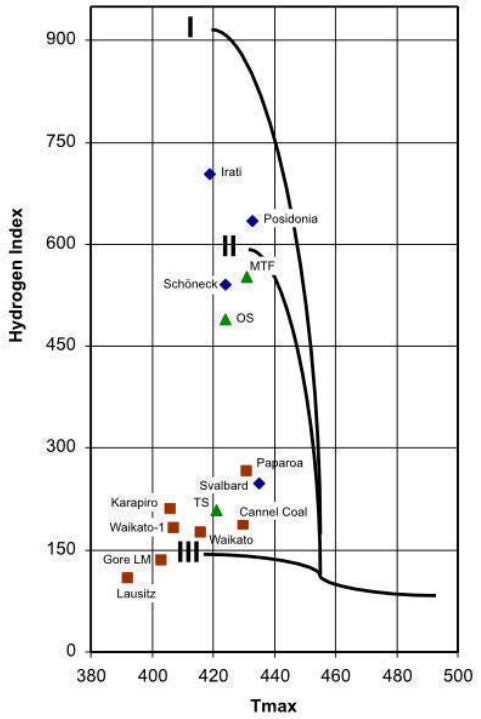


Fig. 2.1 Hydrogen Index vs Tmax of samples. HI, mg-HC/g-TOC; OI, mg-CO2/gTOC.

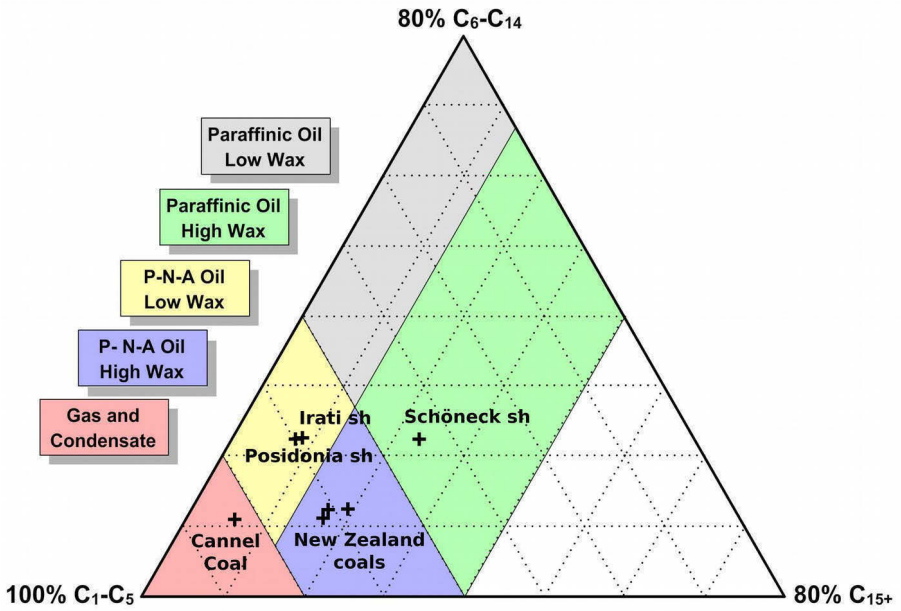


Fig. 2.2 Petroleum type organofacies for some samples.
Source: Mahlstedt (2012) and Vu (2008).

Schöneck shale (G004070)

The Molasse Basin was formed in the Tertiary and comprises of 4 elements, namely (from bottom) Schöneck Fm, previously called as Fish Shale; Dynow Fm; Eggerding Fm and Zupfing Fm. This sample belongs to Schöneck Formation and was taken from a layer of a carbonate-free black shale, it contains 10.1 %wt TOC, a calculated R_O of 0.47 % and Tmax of 424 °C. The presence of pyrite indicates an anoxic depositional environment. This is also supported by low pristane/phytane (< 1.5) determined from the aliphatic fraction of the shale extract. Kaolinite and illite are major mineral components of this shale with quite amounts of lamalginite (Schulz et al., 2002). These authors carried out a characterization of source rocks in the Molasse Basin and concluded that the depositional environment during the sedimentation of this sample involved a stratificed water body triggered by an input of saline surface water, denitrification and CO2- recycling within the water column, resulting in dysoxic-anoxic waters within the photic zone. These prevailing conditions suggest a decrease in the biological activity in comparison to the deeper sedimentary units. Mahlstedt, (2012) reported that the pyrolysate of Schöneck shale exhibited the highest proportion of long-chain *n*-alkanes and *n*-alkenes, which have been related to well preserved outer cell walls of microalgae. This shale is, thus classified as kerogen of paraffinic high wax oil. As expected for a marine shale, the phenol fraction in this sample was low.

Alum shale (G008513)

As a source of hydrocarbons, Alum shale from Bornholm (Denmark) has been subject to several studies. Pedersen (1989) reported the sedimentary facies of two wells drilled in 1984, namely Skelbro-1 and Billegrav-1. The G008513 sample was collected from Skelbro-2 well, which was drilled in 2010 as part of Gas Shale in Europe (GASH) project. Schovsbo *et at.* (2011) reported that the cored succession was significantly similar to the previously described Skelbro-1 well. G008513 is a black shale from Alum Shale Formation with a TOC content of 11.4%. Dated as Middle Cambrian to Lower Ordovician, the organic matter of Alum shale is free of terrestrial macerals (algal-type origin), thus the maturity of the core has been calculated based on vitrinite-like particles, and its value is 2.3 % Ro.

Alum Shale exhibits beds and nodules of limestone deposited slowly under anoxic conditions, this facies also generally lacked fossils (Pedersen, 1989). Mineralogical investigations indicate 50-65% quartz, 5-15% of feldspars and 10-15% muscovite and barite, and ca. 11% pyrite. Barite and pyrite occur as disseminated crystals and as nodules (GASH, 2011). Jensenius (1987) studied the diagenesis of Lower Palaezoic sediments through fluid inclusions and showed that the area of Bornholm could have attained temperatures around 200-220°C. Rock-Eval pyrolysis analysis also indicates high thermal alteration, i.e. a Tmax of 597 °C.

Table 2.1 Formations, ages, depositional environment and Rock-Eval Pyrolysis data of the samples.

Geological setting							Rock-Eval Pyrolysis								
Code	**Name**	**Formation**	**Basin**	**Country**	**Age**	**Depositional environment**	**Kerogen type**	**TOC**	**S1**	**S2**	**S3**	**HI**	**OI**	**Tmax**	**Calc. R_0 *R_0**
G001799	Botneheia	Botneheia Formation		Norway	Middle Triassic	Marine Cenozoic	II/III	2,17	1,5	5,4		249		435	*0.41
G004070	Schöneck	Schöneck Fm.		Austria	Eocene-Oligocene	Marine	II	10,1	2,34	54,62		540		424	0,47
G008513	Alum	Alum Shale		Denmark	Cambrian-Furongian	Marine	II	11,2	0,04	0,89	1,26	8	11	597	2,3
G007152	Posidonia	Posidonia Shale		Germany	Toarcian	Marine	II	11,4	2,97	72,34	2,25	635	20	433	*0.53
G005812	Irati	Irati Fm.		Brazil	Permian	Marine paleozoic	I	25,3	9,68	177,87	1,55	703	6	419	0,39
G007923	Karapiro	Karapiro	Waikato	New Zealand	Pleistocene	Fluvio-deltaic terrestrial	III	45,38	10,72	95,36	40,34	211	57	406	*0.29
G001975	Gore LM	Gore Lignite Measures	Eastern Southland	New Zealand	Oligocene-Miocene	Fluvio-deltaic terrestrial	III	51,61	9,8	69,6	39,5	135	77	403	*0.29
G004548	Waikato-1	Waikato Coal Measures	Waikato	New Zealand	Late Eocene	Fluvio-deltaic terrestrial	III	60,41	2,2	110,26	15,91	183	37	407	0,17
G001992	Waikato	Waikato Coal Measures	Waikato	New Zealand	Eocene-Oligocene	Fluvio-deltaic terrestrial	III	64,92	4,5	113,9	20,5	176	32	416	*0.49
G000698	Cannel Coal	Cannel Coal			Mississipian	Mixed lacustrine - terrestrial	I/III	67,9	1,65	127,27	5,14	187	8	430	0,58
G008814	Lausitz	Brieske	NWET	Germany	Miocene		III	69,4	4,12	75,82	44,99	109	65	392	*0.57
G001990	Paparoa	Paparoa Coal Measures	West Coast	New Zealand	Late Cretaceous	Fluvio-deltaic terrestrial	III	74,18	7,6	198,3	4,3	267	6	431	*0.71
G009482	Oil sand (OS)		WCSB	Canada			II	12	34,12	58,76	3,65	490	30	424	0,472
G011096	Matured Fine Tailing (MFT)		WCSB	Canada			I/II	15,4	12,08	84,99	1,78	552	12	431	0,598
G011107	Tailings sand (TS)		WCSB	Canada			III	3,75	1,89	7,8	7,76	208	207	421	0,418

Source of data: Weiss *et al.* (2000); Schulz *et al.* (2002); Vu (2008); Mahlstedt (2012); Noah *et al.* (2014). Calc R_0, calculated Vitrinite Reflectance according Jarvie *et al.* (2001), Calc. R_0 = 0.018*Tmax – 7.16; *R_0, measured Vitrinite (%); S1, S2; mgHC/g rock; S3, mgCO2/g rock; HI; mgHC/g TOC; OI, mgCO2/g TOC; Tmax, °C; NEWT, North-West European Tertiary Basin; WCSB, West Canada Sedimentary basin.

Posidonia Shale (G007152)

Several studies have been carried out on the Posidonia Shale (Lower Toarcian) as a hydrocarbon source and due to its unique geological conditions, e.g. wide maturity range with low organic facies variations within few kilometers. These sediments were deposited in an inland sea deep enough to develop sub-anoxic conditions in the bottom waters, with limited connection to the open sea (Rullkötter *et al*., 1988). The sample included in the present study (TOC 11.4 %wt and R_0 of 0.53 %), is a marine algal type II kerogen taken from the Wickensen well in the Hils syncline (northern Germany) (GASH, 2010), the units of which are medium — olive gray, fine-grained, well laminated, rich in organic matter and calcite (Littke et al., 1991). The same author described a threefold stratigraphic division for Posidonia Shale, namely lower marlstone, middle calcareous clay-shale with bivalve shells, and upper calcareous clayshale. Although lithological differences can be considered to be relatively small, a difference between the marlstone and shales is the carbonate content which occurs as very thin, white, chalky disks derived from coccolithophoridae and schizospheres (Littke et al., 1991). Recently, Bernard *et al*., (2012) reported that organic matter of Wickensen well contains up to 90% alginites and liptodetrinites (lesser amount), with traces of vitrinite and inertinite. Calcite composes of coccoliths and other plankton-derived microfossils. Framboidal pyrite was also identified in Posidonia shale by these authors.

TEM and STXM analysis showed high heterogeneity in the Wickensen samples at nanometer scale, e.g. it contained biogenic calcium carbonates, and phosphates with detrital quartz. Further, no significant nanoporosity was observed in Wickensen shales, the total porosity of the matrix was filled with homogeneous organic material containing sulphur moieties; this sulphur incorporation into the organic matter is likely related to bacterial sulphate reduction occurring at the beginning of sedimentation (Bernard *et al*., 2012).

Irati Shale (G005812)

Irati Formation is located in Paraná Basin (Brazil). The shale is dark gray, brown, and

black, very fine grained and finely laminated with mica and clay mineral content ranging between 60 to 70 % (Padula, 1969). Mahlstedt and (2012) reported this sample as a Permian shale with a TOC of 25.3 %, having a marine origin, a calculated R_0 of 0.39 % and type II kerogen, although it could also be considered as type I given its high HI (703 mgHC/g TOC). Different depositional environments have been proposed for the Irati formation. Some models propose a low salinity marine basin Padula, (1969), while others infer a lacustrine enviroment (Correa Da Silva and Cornford, 1985). In this organic-rich sediments palynomorphs, plants, reptile skeletons and crustaceans have also been found (Holz et al., 2010).

Irati and Posidonia shales can be classified as P-N-A Oil Low Wax due to higher presence of intermediate normal alkyl-chained compounds and an increase of aromatic compounds (e.g. alkylbenzenes). It is accepted that this organic matter originates from algal and bacterial material, and terrestrial debris (Mahlstedt, 2012).

New Zealand coals

Three lignite samples were chosen from the Waikato Basin, one sample from the Karapiro Formation (G007923) and two from the Waikato Formation (G004548 and G001992). Karapiro (G007923) and Waikato-1 (G004548) samples were taken from the DEBITS-1 well drilled in the Ohinewai coal field during a campaign in 2004 (Kallmeyer *et al*., 2006), while the Waikato (G001992) is from the Rotowaro coal field. Coals from the Gore Lignite Measures Fm (Gore LM, G001975), the Eastern Southland Basin, and Paparoa Coal Measures (G001990) from the West Coast Basin were also included. These coals, ranging in age from late Cretaceous (Paparoa) up to Pleistocene (Karapiro), have variable TOC (45-74 % wt) and Ro (0.29-0.71 %). Vu, (2008) reported Waikato, Gore LM and Paparoa coals as III type kerogen that originated from terrestrial plant material deposited in an oxidizing environment.

The geological record in Waikato Basin involves two main phases: early Cretaceous Rangitata Orogeny producing a strong deformation of the Mesozoic strata and uplift of the New Zealand continental block; and the Neogene phase of deformation (Kaikoura Orogeny) which resulted in block faulting and tilting, thus affecting the basement and Paleogene to early Neogene cover (Te Kuiti, Waitemata, Mahoenui, and Mohakatino groups) (Vu, 2008). Among the lithostratigraphic groups that

comprise the Waikato Basin, G007923 corresponds to Tauranga Group, while G004548 and G001992 belong to the Te Kuiti Group. Tauranga Group sediments have been dated as Miocene and include sediments of terrestrial origin with embedded lignite layers in pumiceous and rhyolitic sands, clays and gravels (Glombitza et al., 2009). The Te Kuiti Group, which underlies the Tauranga Gp, incorporates Late Eocene and Early Miocene sediments in which carbonate content increases towards upper layers. These sediments are also thought to have been deposited at an estimated temperature of up 75 °C (Vieth et al., 2008).

The West Coast Basin region of the South Island of New Zealand occupies the eastern edge of the Australian plate. It originated from Gondwana prior to the Cretaceous, displaying periods of terrestrial-sediment deposition and subsidences which resulted in marine sandstones, mudstones and limestones. It also underwent tectonic deformation close to Late Oligocene and Early Miocene. The West Coast basin contains alternating layers of sandstones, coals and lacustrine mudstones. Sampled from the West Coast, the G001990 coal belongs to Paparoa Coal Measures Fm (Late Cretaceous) (Vu, 2008).

Eastern Southland Basin comprises an Oligocene - Miocene sedimentary sequence which includes several lignite seams within marine and non-marine facies. The Gore Lignite Measures dominates the Eastern Southland Basin and contains conglomeratic deposits and coarse sandstones interbedded with lignite. The G001975 coal was sampled from this Formation (Vu, 2008).

All the three coals, namely Gore LM, Waikato and Paparoa, have been described as P-N-A high wax kerogen (Horsfield, 1989). Given that lignin/cellulose-derivative products are characteristic components of the coals, when laboratory pyrolysis is applied, the pyrolysates are richer in phenolic compounds, long chain n-alkanes and naphthenoaromatic compounds than in pyrolysates of shales (Mahlstedt, 2012).

Cannel Coal (G000698) and Lausitz Coal (G008814)

G008814 is a German coal from the Lausitz area (Brandenburg, open-cast mining "Welzow-Süd") with a TOC of 69.4 %, a Tmax of 392 °C (Noah M, unpublished). This Lausitz coal belongs to the Brieske Fm, which lies between the Rauno Fm (below)

and Spremberg Fm (above). The area has undergone several periods of compression, subsidence and sedimentation, along with a series of long-term and short-term transgressions and regressions. Facies in the Brieske Fm show vertical and lateral variability, the former likely due to changes of water level, and the latter related to nutrient content derived from terrestrial input (Seifert et al. 1993; Standke et al. 1993).

Mahlstedt, (2012) reported G000698 as a Mississipian sapropelic coal with a TOC content of 67.9 %, Tmax of 430 °C and I/III kerogen type, which was formed under a mixed lacustrine-terrestrial depositional environment. The author concluded that kerogen was gas-prone and the pyrolysate was characterized by the predominance of phenol and aromatic compounds, additionally it exhibited high abundance of short and long-chained aliphatic components due to a sporal organic matter.

Although limited information about the origin of the coal samples is available, these are valuable as they complete the overview on TOC rich samples of variable origin and geologic age and their feedstock potential for the deep biosphere.

Oil-sand related samples

The set included three sediments from three of the major stages of the processing and reclamation procedure: an oil sand (OS) (G009482) from an open-pit mine, a processed mature fine tailing (MTF) (G011096) from a drying cell, and a tailings sand (TS) (G011107) from a reclamation site (Fig. 2.3).

Fig. 2.3 Appearance of processed mature fine tailing (a, b) and tailings sand (c, d) of oil-sand related samples from Alberta, Canada.

By definition an oil sand is a mixture of bitumen, sand and clay. Rock-Eval pyrolysis analysis applied to this sample suggests a kerogen type II (Fig. 2.1), with a TOC content of 12 % wt. A Tmax of 392 °C suggests that this sample can be considered as immature. For the MFT, the TOC was 15.4 %wt with a Tmax of 431 °C. This sample still contains a portion of bitumen and has been conditioned with polyacrylamide to reduce the water content. The tailings sand has a TOC of 3.75 %wt and a Tmax of 421 °C. This sample has also been treated by addition of peat-mineral mix and exhibited plant growth (Fig 2.3c, d).

Noah *et al*. (2014) accomplished a biochemical characterization of the sediments, from the mining phase up to recultivation process. A screening of the compound classes showed that only TS exhibited presence of *n*-alkanes, i.e. this fraction was absent in OS and MTF samples; in contrast, naphtalenes and chrysene were not found in TS. In all the samples, the rest of compound classes (hopanes, steranes, phenanthrenes, benz[*a*]anthracene, aromatic steroids and fatty acids) were present. The authors also observed that the organic content in MFT is related to clay content, higher clay fraction correlates with higher TOC, likely because clays tend to retain more non-recovered bitumen. Biomarkers such as terpanes, C35-homohopane Index and aromatic steroids have shown that biological processes or heating in drying cells or tailings ponds may cause no significant changes in the composition of the organic

matter; however, the amount of some compounds (e.g. naphtalenes) may vary due to ongoing microbial metabolism. Further, cell counting proves that the oil-extraction process has no effect on cell abundances. It is evident that the original composition has been also affected by the addition of dewatering agents and peat-mineral mix, either due to their properties themselves or by stimulation of biological processes (e.g. plant development).

3. Methodology

This section describes the techniques employed to characterize the LMWOA released from organic-rich sediments. Preliminary description of the samples was made by means of Rock-Eval and TOC analyses. Afterwards, finely crushed sediments underwent an aqueous extraction as a laboratory equivalent of natural processes in deep environments that releases LMWOA. Consequently, the aqueous extracts has been characterized regarding dissolved organic carbon (DOC) and its fractions by LC-OCD; and underivatized water-soluble carboxylic acids by ion chromatography (IC) and solid-phase extraction (SPE). The organic SPE extract is analyzed by gas chromatography with flame ion detection (GC-FID) and gas chromatography-mass spectrometry (Fig. 3.1). IC and SPE-GC-FID use external calibration for quantification. The description below begins with the description of LC-OCD analysis, followed by IC procedure and the developed SPE method.

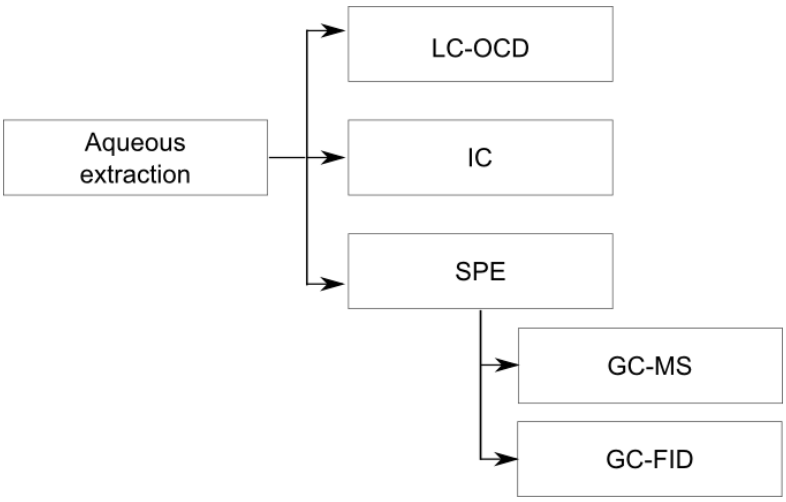


Fig. 3.1 General procedure for characterization of aqueous extracts

3.1 TOC and Rock-Eval

Samples were provided from different projects, as were TOC content and Rock-Eval pyrolysis data. Such geochemical methods have been carried out following standards procedures, briefly described in the following paragraphs.

Total organic carbon (TOC) was determined using a LECO analyzer, model TM -CNS-

200 or SC-632 (Vu, 2008; Mahlstedt, 2012). As first step inorganic carbon is removed from the sample adding diluted HCl until it doesn't exhibit effervescence. Secondly combustion is applied in temperature increments up to over 1000°C. An infrared-based system translates the variation of energy during the heating into carbon content.

Rock-Eval pyrolysis (Espitalié et al., 1977) was carried out by a Rock-Eval Instrument 6. The standard procedure includes heating the sample iso- and no-isothermally up to 650 °C. The combustion products, analyzed using a flame ionization detector in a helium gas flow, describe three significant peaks. The first peak, S1, represents free hydrocarbons and relates to the present maturity level of the organic matter; the second peak, S2, represents the products formed by thermal cracking; finally, S3 relates to CO_2 produced during the cooling of the pyrolysis oven, which has been interpreted as a measure for the oxygen richness of the kerogen. These three values, S1, S2 and S3, along with TOC content, serve to calculate the HI-Hydrogen Index, OI-Oxygen Index, and PI-Production Index of a sample, Tmax is the maximum temperature of S2 peak.

3.2 Simulation of LMWOA release: aqueous extraction

The extraction was carried out using an assembled apparatus that consisted of a two-neck angled round-bottom flask and a condenser, the heat was provided by a hot plate (Fig 3.2). 250 ml of water were placed in the flask together with 5 or 10 g of sediment and some boiling stones (to prevent ebullition), and warmed up to 80 C° - 90 °C, for 72 h. The amount of sediment employed in the extraction depended on sample TOC content, it was 10 g for samples with TOC below 25%, and 5 g for samples with higher TOC content. After 72 h, the particles larger than 45µm were removed from the extract by means of a vacuum line and a filtration pump; with polypropylene filters. Secondly the aqueous extract was put in a 250-ml volumetric flask, which was filled up to the mark with millipore water to 250 ml volume following which, 25 ml were taken out for IC chromatography and LC-OCD analyses. The resting 225 ml were acidified with 5 M HCl up to a pH below 2, ca. 0.5 ml per 50 ml of extract, and millipore water was added to get a 250-ml solution. Subsequently the extracts were stored in cold (ca. 3 °C) and

darkness until the next procedures.

Fig. 3.2 Water extraction

3.2.1 Extraction in different water/sediment ratios

The tested water/sediment ratios were 500:1, 100:1 and 50:1, corresponding to 0.5, 2.5 and 5 g of sediment respectively. Extraction was done in 250 ml of water and the selected amount of sediment, keeping temperature and heating time conditions as described above, i.e. at 80 C° - 90 °C for 72 h.

3.2.2 Extraction with phosphate buffers at different pH

For testing the effect of pH and phosphate content, the extractions were done with 5 g of sediments in aqueous buffer solutions containing conjugate acid-base pairs shown in Table 3.1. For the solution with pH 2, phosphoric acid and monopotasium phosphate; for the solution with pH 7, monopotassium phosphate and dipotassium phosphate; and for the solution with pH 12.5, dipotassium phosphate and tripotassium phosphate.

Table 3.1 Buffer solution formulas

Acid, conjugate acid	Base, conjugate base	pH value
Phosphoric acid (85%) 6.8 ml	Monopotassium phosphate 12.8 g	2
Monopotassium phosphate 12.8 g	Dipotassium phosphate 15.8 g	7
Dipotassium phosphate 15.8 g	Tripotassium phosphate 19.6 g	12.5

*1 l solution volume, pH was adjusted using a pH meter and addition of either 6 M HCl or 6 M NaOH.

3.3 Size-exclusion chromatography (LC-OCD)

DOC and its fractions were characterized by means of gel permeation chromatography, or size-exclusion-chromatography, which assembles organic carbon detection (IR) and UV detection (λ=254 nm) (Huber & Frimmel, 1996). The mobile phase employed was phosphate buffer (pH 6.6; 9 mM Na_2HPO_4, 18 mM KH_2PO_4), with a flow rate of 1 ml/min. As a preparatory step to the chromatographic column (Novo-Grom GP2, 300 mm x 8 mm, Alltech Grom GmbH) the sample was filtered using a 0.45 µm membrane. Once DOC has been divided into fractions, namely biopolymers, humic substances, building blocks, acid- and neutral-LMW compounds, they were characterized by UV detection. Quantification is made by UV oxidation (λ=185 nm) in a Gräntzel thin-film reactor by IR-detection. To calculate the proportion of aromatic structures, a ratio between the spectral absorption coefficient (SAC in m -1 at 254 nm) and the DOC (in mg C/L) (an indicator used for humic substances) has been used. This procedure relies on the distribution of molecular weight; thus, for calibration the standards used were humic and fulvic acid of the Suwannee River.

3.4 Ion chromatography (IC)

For determination of the inorganic anions, aqueous direct injection onto an ICS 3000 chromatograph system (Dionex Corp) was employed. This instrument integrates a conductivity detector, a KOH eluent generator and ASRS Ultra II 2 mm suppressor, and a column AS11HC (2 x 250 mm) with an AG11HC pre-column. The operational conditions are those established by Vieth *et al.* (2008), briefly samples are eluted using a KOH solution, with concentrations ranging from 0.5 mM up to 100 mM at 35 °C, and an external calibration was used for quantification.

3.5 Solid phase extraction (SPE)

The SPE apparatus was built using a syringe pump, a 200-µl loop valve, three 3-way switching valves and a packed stainless steel column, with an inner Ø of 4 mm and 40 mm length (Fig. 3.3). The SPE column was filled with two reversed solid phases Supelclean ENVI-18 and LiChrolut® EN in a ratio of 1:1, ca. 40 mg each, and inert sand, and then conditioned with 1 ml of mixture of methanol-acetonitrile and subsequently with 1 ml of water.

Fig. 3.3 Semiautomatic SPE apparatus.

The procedure of SPE starts with the acidification of the aqueous extract with 5M HCl to pH below 2, 0.5 ml HCl per 50 ml-sample. Subsequently, 50 ml of the acidified aqueous extract is injected onto the column with a flow rate of 2 ml/min, followed by 30 ml of air at a flow of 3 ml/min, to remove the water excess of the column; the direction of the flow is then inverted and 200 µl of millipore water are injected via the loop valve, to eliminate the HCl remains along with 50 ml of air blown at 3 ml/min to dry the column. The elution of analytes is done with 200 µl of acidified DMF (10% acetic acid) at 3 ml/min (Fig. 3.4). SPE extracts were analyzed by GC-MS for qualification and by GC-FID for quantification.

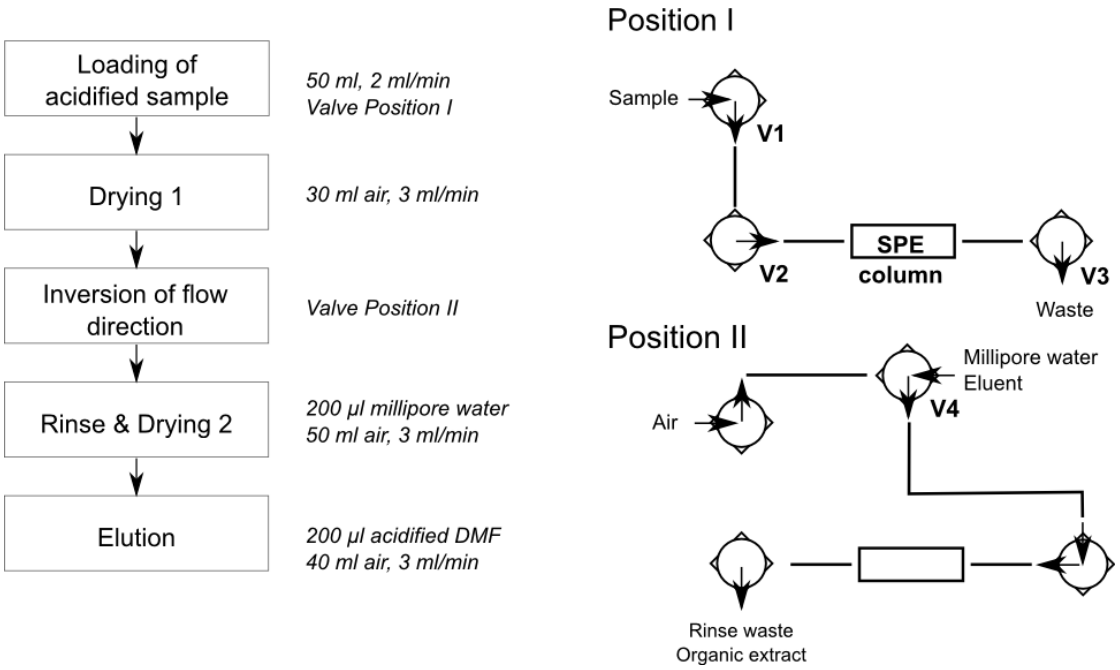


Fig. 3.4 SPE procedure

3.5.1 External calibration of SPE-GC-FID

Calibration curves were made by applying SPE to a series of 10 solutions by duplicate. The aqueous standard solutions contained 10 acids of known concentrations, 2.5, 5, 10, 15, 20, 40, 60, 80, 100 and 120 µg/l. The SPE procedure was applied to acidified Millipore water by triplicate, consequently, it was applied to the aqueous standard

solutions starting from the lowest to the highest concentration one time. Afterwards, before the second set was applied, a run was carried out using acidified water. The SPE extracts were analyzed by GC-FID in triplicate. In order to calculate the calibration curves, Mandel's fitting test was employed to set the linear range. Curves were calculated for propanoic, 2-methylpropanoic, butanoic, pentanoic, hexanoic, heptanoic, octanoic, cyclohexanecarboxylic, nonanoic, decanoic acid, and 4-ethyloctanoic acid served as internal standard.

3.6 Gas chromatography

SPE extracts were analyzed by GC-FID and GC-MS for quantification and compound identification respectively. The chromatographs were equipped with an INNOWax column and operated under the following conditions: the injector temperature was at 40°C and it increased at a rate of 12 °C/s to 250°C, then kept constant at that temperature for 10 minutes. The oven temperature started at 40 °C kept during 2.5 min and then increased at a rate of 10 °C/min up to 200 °C and 8 °C/min to 250 °C during 10 min. Helium, set to a flow rate of 1.0 ml/min, was used as carrier gas. GC-FID detector was set at 300 °C. For compound identification the MS was operating in electron impact mode at 70 eV.

4. Methodological development

One of the objectives of the present study was to develop a method to characterize underivatized LMWOA containing between one to 9 carbon atoms that are released from organic-rich sediments when exposed to heat and water - conditions commonly present in subsurface environment. In this study, two different acid-base LLE procedures were tested, but due to low recovery both were discarded. Subsequently, a SPE procedure was adapted as sample preparation method. GC-MS and GC-FID, were employed as analytical technique. The methodological approach and the modifications that should be carried out due to a series of unexpected results are described.

4.1 Calibration and sensitivity in GC-FID and GC-MS

Gas chromatography is often used to analyze organic acids and **one-point calibration** is usually employed to calculate their concentrations. For this calibration a compound in known concentration, namely the internal standard (IS), is added to the sample or sample extract. Subsequently, the analyte concentration can be calculated from the IS peak area, its concentration and the peak area of analyte. Therefore it must be assumed that the signal response and peak shape of the analyte is similar to the internal standard (Hibbert 2007). In accordance with the International Vocabulary of Metrology (VIM 2012), “sensitivity” is the quotient of the change in the indication and the corresponding change in the value of the quantity being measured. Mathematically, sensitivity is the slope of the function of the peak area related to the concentration. Fig. 4.1 shows that for compound I the sensitivity is higher than for compound III, i.e. the changes on peak area due to concentration variations are larger for I in comparison with the ones for III. Thus, the suitability of one-point calibration must be determined in case of underivatized LMWOA.

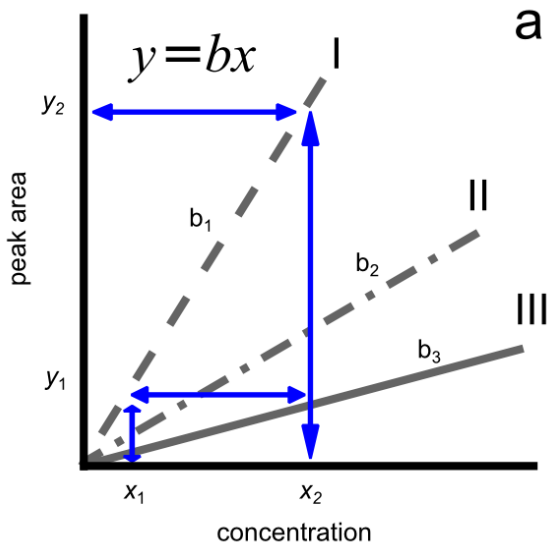


Fig. 4.1 A schematic diagram describing sensitivity. The slope ***m*** of the line ***y*** represents the sensitivity.

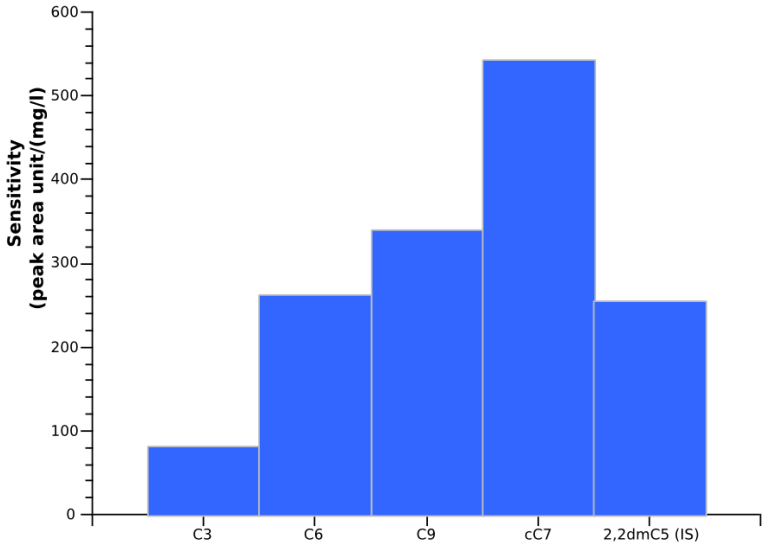


Fig. 4.2 Differences in sensitivity for five carboxylic acids.

For the first acid-base approach LLE, the sensitivity using GC-MS was determined for the propanoic, hexanoic, nonanoic, cycloheptanecarboxylic and 2,2-dimethylpentanoic acid (Fig. 4.2). Sensitivities were calculated under a linear calibration curve with the origin at zero, standard solutions were made using DCM as solvent, in a concentration range from ca. 0.5 mg/l to 15 mg/l. GC-MS analyses were carried out with a FFAP column on a Finnigan MAT 95 XL chromatograph, in total ion scan mode. Cycloheptanecarboxylic acid showed the strongest sensitivity and propanoic acid the

lowest one. Thus sensitivity strongly differed (e.g. a five-fold difference) amongst these acids.

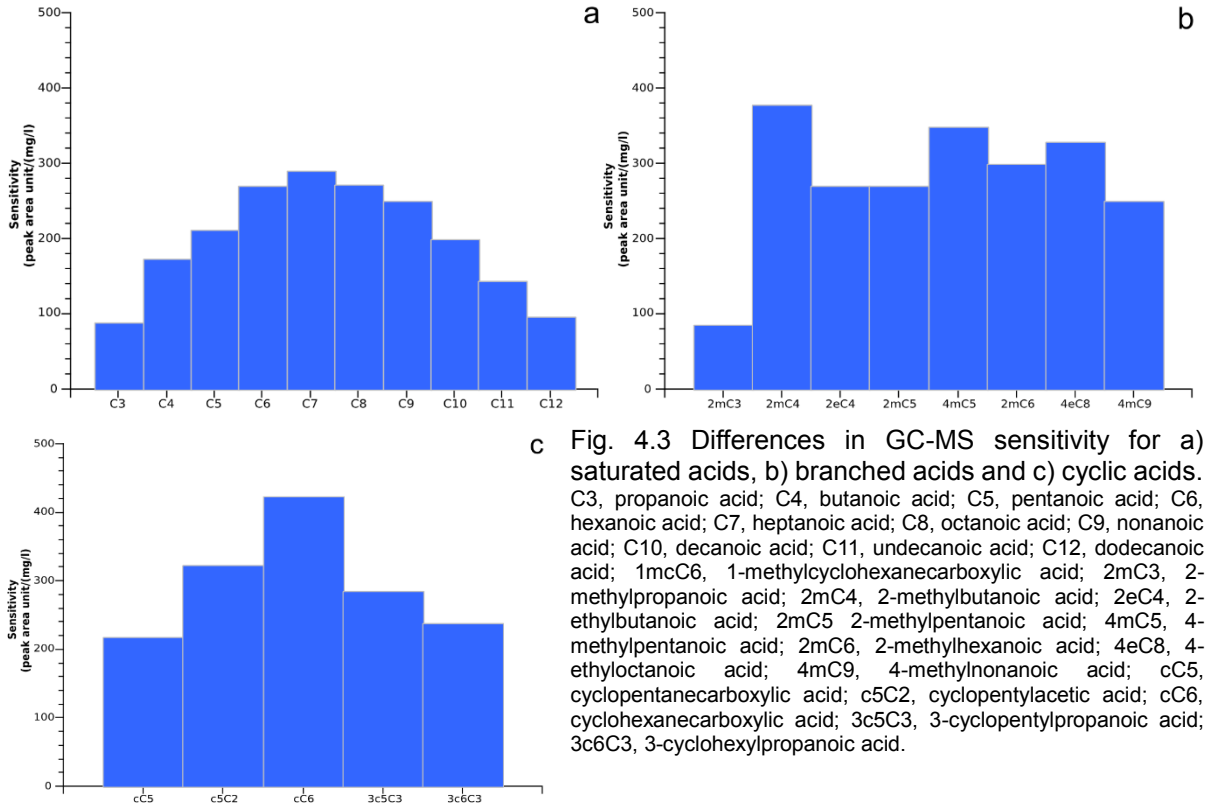


Fig. 4.3 Differences in GC-MS sensitivity for a) saturated acids, b) branched acids and c) cyclic acids. C3, propanoic acid; C4, butanoic acid; C5, pentanoic acid; C6, hexanoic acid; C7, heptanoic acid; C8, octanoic acid; C9, nonanoic acid; C10, decanoic acid; C11, undecanoic acid; C12, dodecanoic acid; 1mcC6, 1-methylcyclohexanecarboxylic acid; 2mC3, 2-methylpropanoic acid; 2mC4, 2-methylbutanoic acid; 2eC4, 2-ethylbutanoic acid; 2mC5 2-methylpentanoic acid; 4mC5, 4-methylpentanoic acid; 2mC6, 2-methylhexanoic acid; 4eC8, 4-ethyloctanoic acid; 4mC9, 4-methylnonanoic acid; cC5, cyclopentanecarboxylic acid; c5C2, cyclopentylacetic acid; cC6, cyclohexanecarboxylic acid; 3c5C3, 3-cyclopentylpropanoic acid; 3c6C3, 3-cyclohexylpropanoic acid.

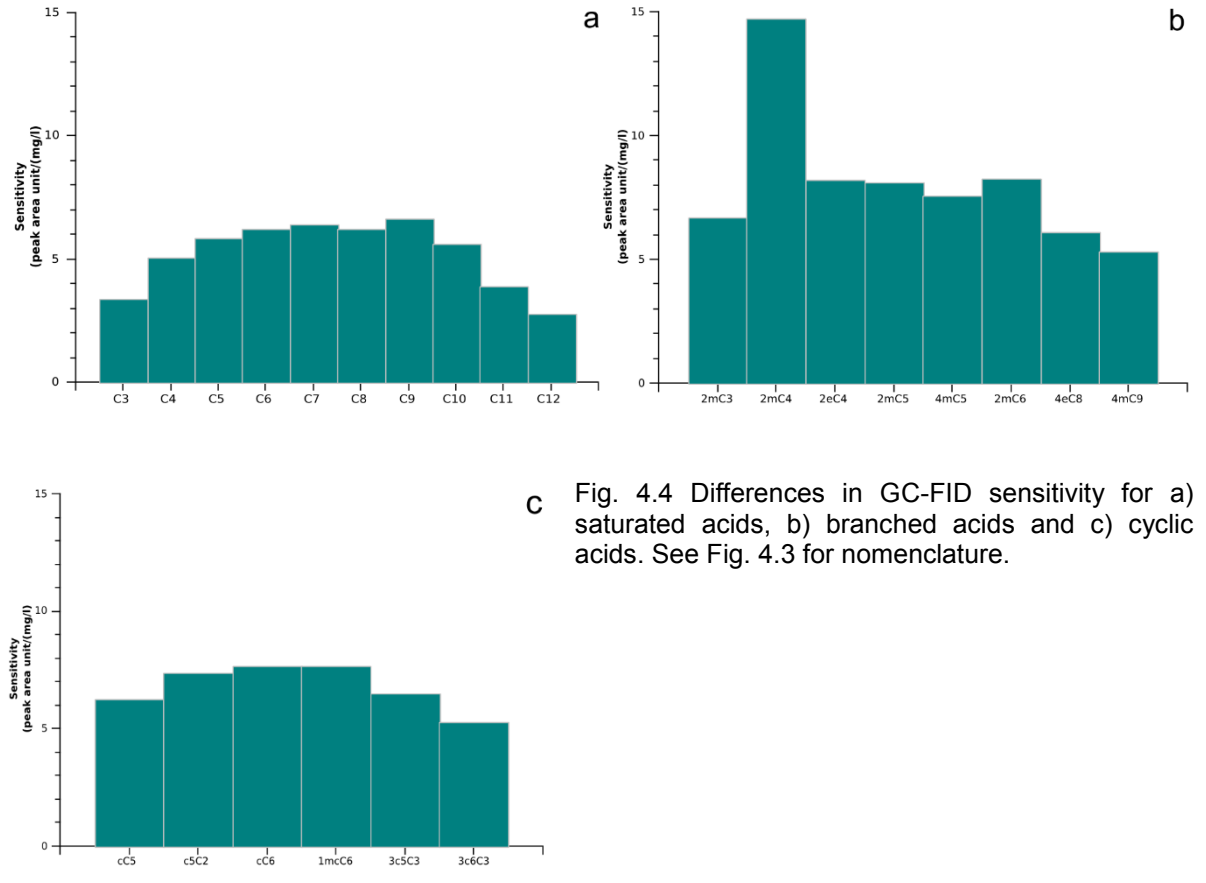


Fig. 4.4 Differences in GC-FID sensitivity for a) saturated acids, b) branched acids and c) cyclic acids. See Fig. 4.3 for nomenclature.

To get an impression about how other potential analytes may behave during analysis, the sensitivity for several types of fatty acids (saturated, branched and cyclic acids) were systematically investigated. Fig. 4.3 shows that for saturated carboxylic acids, sensitivity clearly increases with carbon number length (up to heptanoic acid) and then decreases, i.e. the sensitivity of pentanoic acid is lower than that of hexanoic acid, which in turn is lower than that of heptanoic acid. In the series of saturated acids, propanoic acid exhibited the lowest sensitivity. In the branched acids group, the sensitivity of 2-methylpropanoic acid was lowest for GC-MS, whereas cyclopentanecarboxylic acid was the lowest sensitivity among the compounds in the cyclic group. In general, the three groups of the acids exhibited differences in sensitivity. The variability in sensitivity was also observed in the acids when analyzed by GC-FID (Fig. 4.4).

Ábalos *et al*., (2000) proposed a coupled SPME-GC-FID method to quantify linear free fatty acids, from C2 up to C7, in wastewater samples. For such a method, they reported response factor of propanoic acid is much lower than that of heptanoic acid. As their goal was to evaluate the performance of different SPME fibers under numerous conditions, they did not report the GC-FID sensitivities.

From the differences in sensitivities in both GC systems, an obvious discrimination in sensitivity of linear carboxylic acids related to carbon number was observed. Since one of the goals of the present work is quantification of saturated carboxylic acids, it is important to examine the possible reason for such variations.

Concerning GC-MS, Mohler *et al.*, (1950) observed that with some exceptions total ionization (relative to *n*-butane) increased with increasing number of carbon atoms in series of hydrocarbons. Harrison *et al*., (1966) reported that for homologous series of ethers, esters, ketones, alkanes, alkenes, alkynes and alkylbenzenes, the total ionization of molecules increases with their cross sections and the chain lengths. In this way, it was well established that the electron impact mass spectra response for a molecule depends on its molecular cross section for electron impact (EI) ionization. Thus, the discrimination may partially be explained by the relationship between the fragmentation pattern of carboxylic acids and the *m/z* range of the GC-MS method.

Each point in the chromatogram, total ion current, corresponds to a specific mass spectrum, and the mass spectrum depends on the range of ions to be measured (Fig. 4.5).

To prevent measurement of artificial, disturbing masses e.g. CO_2^+ (*m/z* 44), Ar+ (*m/z* 40) and others, that lead to highly increased noise and interference with analyte signals, standard analysis scans begin at *m/z* 50. This, of course, affects the response of underivatized LMW carboxylic acids due to the fragment ions that are thus not included in the scan. Fig. 4.6 shows the distribution of *m/z* values below and above *m/z* 50 of eight linear carboxylic acids (from propanoic acid up to decanoic acid), it can be seen that for propanoic acid 48% of ions are smaller than m/z 50, and accordingly its sensitivity is the lowest (Figs. 4.3a). For other acids, the abundance of *m/z* values below 50 decreases (e.g. 21.6 % for pentanoic acid and 14.6 % for decanoic acid) with increasing carbon chain length and consequently reduces the above mentioned effect on their response. As explained lines below, this effect can be solved by the use of relative response factors.

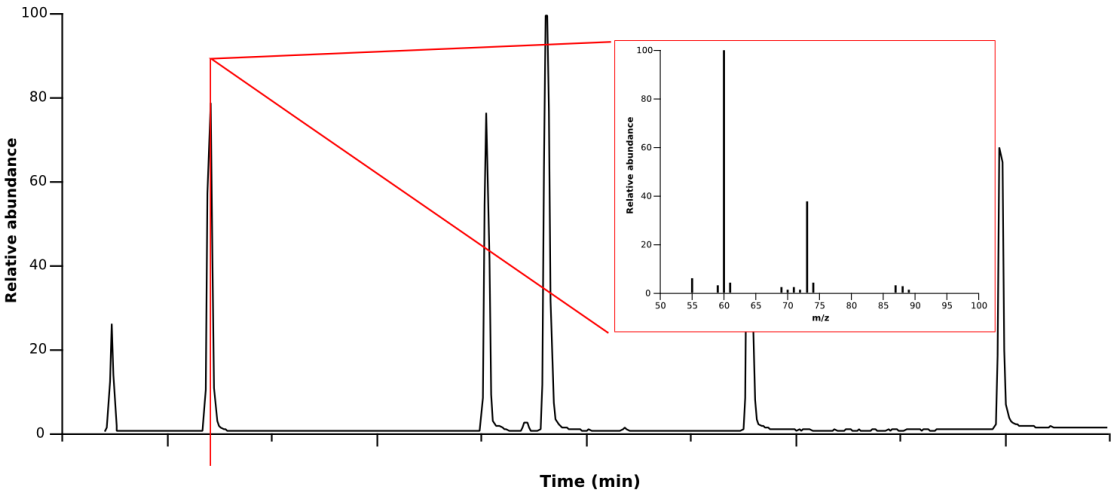


Fig. 4.5 How a chromatogram is built in GC-MS. Each point in the chromatogram, total ion current, corresponds to a specific mass spectrum.

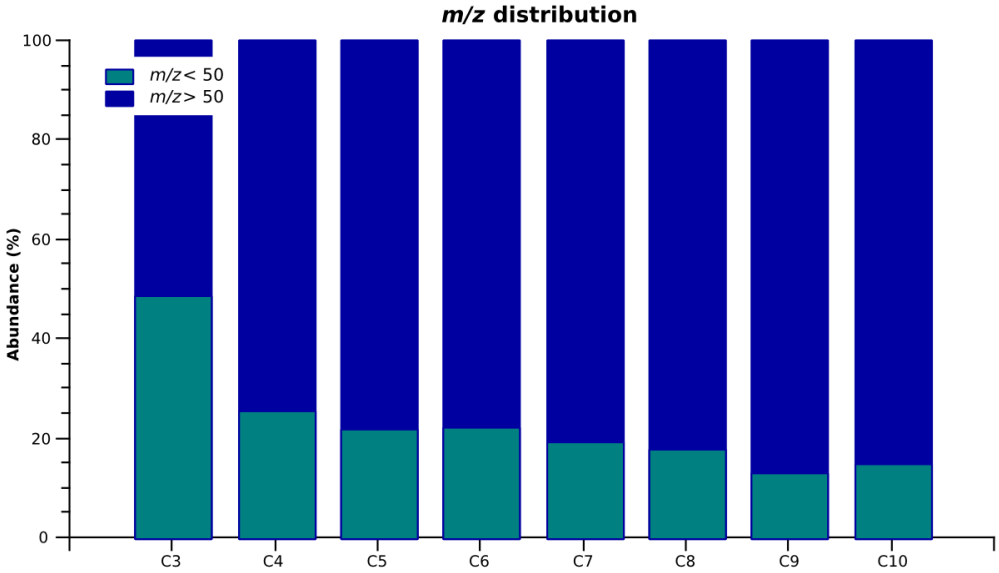


Fig. 4.6 Relative abundance of ions regarding m/z values for a homologous series of n-carboxylic acids. See Fig. 4.3 for nomenclature. Data source: SDBS-AIST, 2015.

Regarding GC-FID, the chromatogram is built with the measurement of the current generated by the collection of ions formed by the flame during the combustion of the compound (McWilliam 1983). For fatty acids with more than six or seven carbon atoms Ackman (1964) reported a linear correlation between response factor and the carbon atom number. It was suggested that this is linked to the proportion of the non-contributing carboxyl group relative to the oxygen-free aliphatic chain (Ackman and Sipos 1964). In general terms, the response of carboxylic acids is the lowest amongst oxygenated compounds and alcohols exhibits the highest. In between these two are ethers, aldehydes, ketones and esters (Scanlon and Willis 1985).

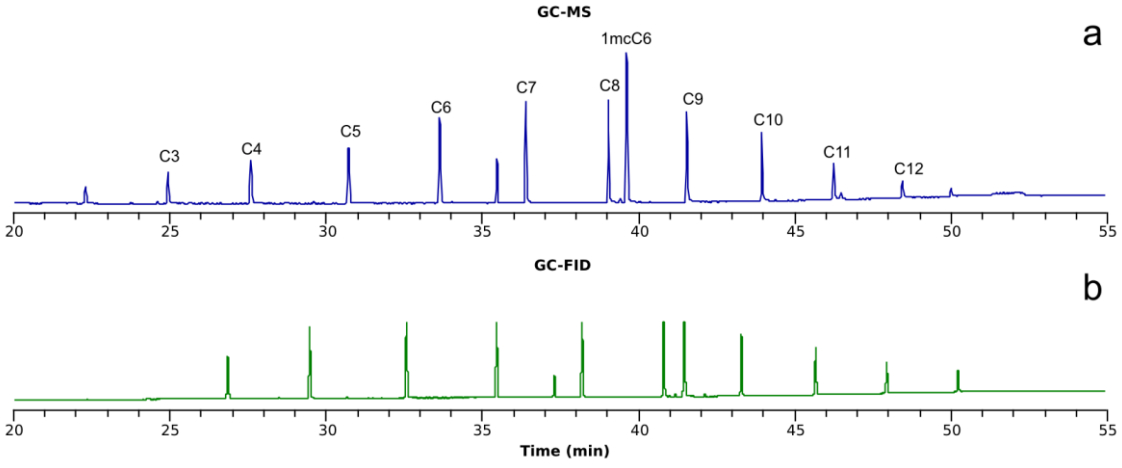


Fig. 4.7 GC-MS chromatogram (blue line) and GC-FID (green line) chromatogram of a standard solution in DCM. See Fig. 4.3 for nomenclature.

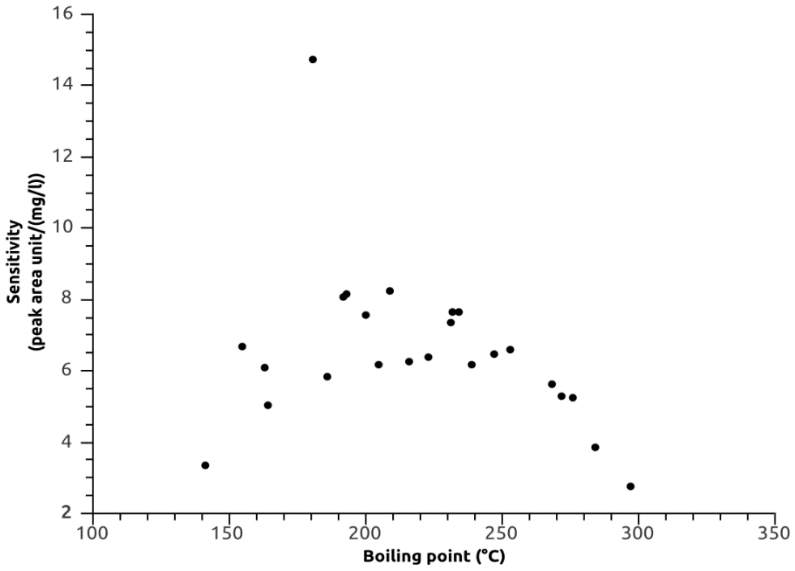


Fig. 4.8 Boiling point vs sensitivity fin GC-FID for selected carboxylic acids. Based on Figs. 4.3 and 4.4, and Table 4.1.

From another standpoint, while looking for an explanation for why acids with carbon chain length longer than 9 atoms, exhibited a low response, either in MS or FID chromatogram of standard solutions (Fig. 4.7), a discrimination between compounds with low and high boiling points was seen (Table 4.1, Fig. 4.8). In gas chromatography the problem of discrimination of high boiling point compounds exists when the injector is not hot enough to vaporize high boilers so they are not effectively transferred to the column (McMaster, 2007a; McMaster, 2007b). Thus, the injector temperature not less than 10 °C below the boiling point of the highest boiling analyte should be chosen.

The boiling point of linear carboxylic acids increases with the carbon chain length (Table 4.1) and reaches about 300 °C for dodecanoic acid – highest value amongst the investigated analytes. However GC columns have an upper temperature limit of operation to avoid thermal damage (FFAP: 250 °C), thus the front inlet should not be higher than 230 °C. The boiling points of octanoic and 1-methylcyclohexanecarboxylic acid are similar, 239 and 234 °C respectively, but there is a significant difference in signal response of such compounds, suggesting that other mechanism (e.g. ionization) may come into consideration as well.

Table 4.1 Boling points of some linear, branched and cyclic carboxylic acids

Linear acids	Boling point (°C)	Branched acids	Boiling point (°C)	Cyclic acids	Boiling point (°C)
Propanoic acid	141	2-Methyl-propanoic acid	155	Cyclopentane-carboxylic acid	216
Butanoic acid	164	2-Methylbutanoic acid	180.3	Cyclopentylacetic acid	231
Pentanoic acid	186	2-Ethylbutanoic acid	193-194	Cyclohexanecarboxylic acid	232-233
Hexanoic acid	205	2-Methyl-pentanoic acid	192-197	1-Methylcyclo-hexanecarboxylic acid	234
Heptanoic acid	223	4-Methyl-pentanoic acid	200	3-Cyclopen-tylpropanoic acid	247.2
Octanoic acid	239	2,2-Dimethylpentanoic acid	207.8	3-Cyclohexylpropanoic acid	275.8
Nonanoic acid	253	2-Methylhexanoic acid	209-210		
Decanoic acid	268-270	4-Ethyloctanoic acid	163		
Undecanoic acid	284	4-methylnonanoic acid	271.6		
Dodecanoic acid	297				

Source: Data from MSDS by Sigma-Aldrich at http://www.sigmaaldrich.com/

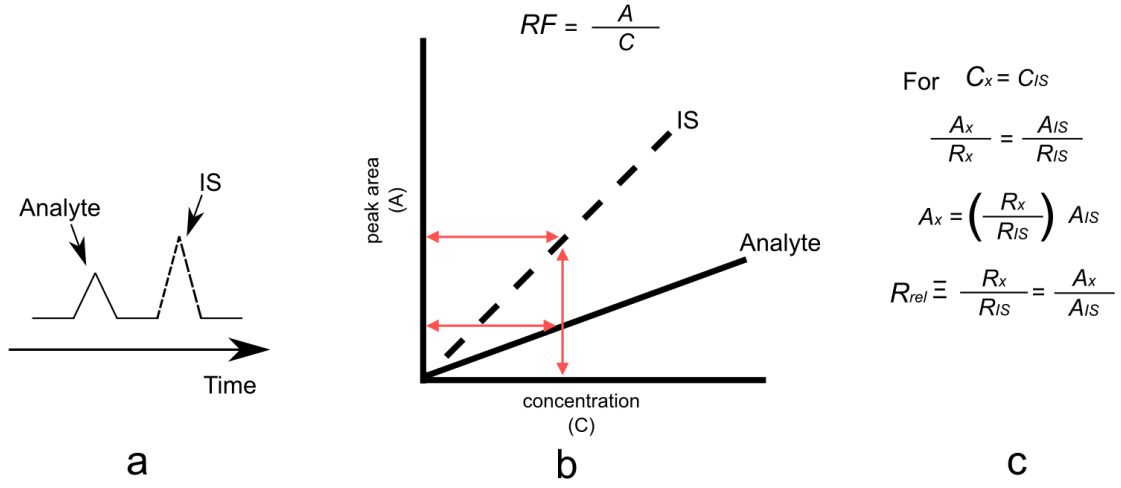


Fig. 4.9 Relative response factors. a) Peak area of internal standard (IS) and analyte (x) are not similar if compared at same concentration. b) Plot that describes the relationship between peak area and concentration with the response factor. c) Mathematical deduction to solve the analyte concentration from the analyte and IS response factors and peak areas.

The observed differences in GC-MS sensitivity amongst the fatty acids demonstrate that the one-point calibration should not be applied. In this study, factors were calculated to compensate such differences, namely relative response factors (see

Cuadros-Rodríguez *et al.*, 2007 for a review). Using standard solutions in DCM or DEE, relative response factors were calculated for individual compounds considering a blank signal equals to zero (Fig. 4.9).

Summarizing, differences in sensitivities in GC-MS and GC-FID amongst various compounds were observed. Particularly in case of linear LMWOA, sensitivity increases up to heptanoic acid and then decreases. The low sensitivities of the compounds on the left half (up to heptanoic acid), are assumed to be related to: i) the exclusion of *m/z* values below 50 in GC-MS; and, ii) the effect of the carboxyl group in comparison to the oxygen-free aliphatic chain. During combustion, the carboxyl carbon doesn't provide ions that are quantified by the detector. The decrease in sensitivity of acids with carbon chain length longer than 9 atoms, is supposed to be due to discrimination of high boilers (e.g. boiling point of nonanoic acid is 253 °C). The front inlet of the chromatograph should not be higher than 230 °C to avoid thermal damage of the column. Due to this differences in sensitivity, the calculation of concentrations should be done using relative response factors.

4.2 Acid-base LLE

Liquid-liquid extraction (LLE), a very popular technique, works on the principle like-dissolves-like. The functional groups relates to the polarity of a compound, and in the case of carboxylic acids, the carboxyl group is highly polar due to the C=O bond and the –OH bond. This solubility of carboxylic acids in water depends on the acid-ionization constant (Ka) and the pH of the solution (Kenkel, 2003). For example, Eq. 4.1 shows the equilibrium equation for pentanoic acid:

$$CH_3(CH_2)_2 - COOH + H_2O \Leftrightarrow H_3O^+ + CH_3(CH_2)_2 - COO^- \quad \text{Eq. 4.1}$$

$CH_3(CH_2)_2$-COOH represents the less soluble (protonated) form of the acid, H_3O^+ stands for hydronium ion and $CH_3(CH_2)_2$-COO^- is the deprotonated (more soluble) acid form. Formic and acetic acid are sufficiently polar to be completely soluble in water, but the solubility of carboxylic acids decreases as the chain length increases.

The distribution of acids between the water and the solvent during LLE is determined by the distribution or partition coefficient which is the ratio of the solubility of analyte dissolved in the organic layer to the solubility of acid dissolved in the aqueous layer.

$$Distribution\ coefficient:\ K = \frac{Solubility\ of\ organic\ solvent\ (g/100ml)}{Solubility\ of\ water\ (g/100ml)} \quad \text{Eq. 4.2}$$

From this equation, it can be seen that just certain amount of the acid will move into the organic solvent in the first extraction, thus it is a common practice to perform successive extractions to enhance the recovery of the target analytes.

4.2.1 First approach

To evaluate the suitability of LLE for the expected complex mixture of compounds released into the aqueous extract of the organic-rich sediments, LLE was applied to the aqueous extracts of four samples covering a wide range of TOC (Table 4.2). Fig. 4.10 shows the design of first LLE approach with dichloromethane (DCM). The organic extracts were analyzed by GC-MS. The aqueous sample was divided into two parts to apply LLE under different conditions. For the first subsample, LLE was performed twice, firstly to the aqueous soxhlet extract to remove non-polar compounds, obtaining the A extract; secondly to the aliquot acidified extracted aqueous solution to recover the acids, resulting in B extract. For the second subsample, the soxhlet extract was acidified directly and extracted with DCM, resulting in the C extract. The acidification will favor protonation of the acids to improve the extraction. Subsequently, the organic extracts were concentrated by means of a Turbovap® down to ca. 1-ml and analyzed by GC-MS on a FFAP column.

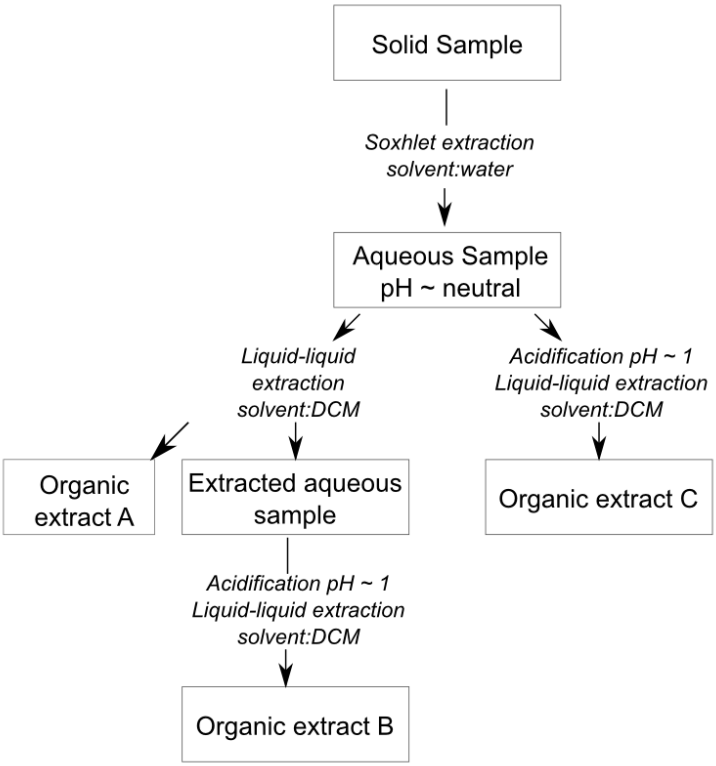


Fig. 4.10 First approach using liquid-liquid extraction (LLE).

Table 4.2 Samples used for the first LLE approach

Sample	TOC (%wt)	Ro (%)	Lithology	Age
G007022	0.1	n.a.	Sandstone	n.a.
G001799	2.2	0.41	Shale	Middle Anisian
G007923	45.5	0.29	Lignite	Pleistocene
G004548	60.4	0.39	Black coal	Late Eocene

GC-MS chromatograms of the different extracts of G001799 show the results of the first LLE approach (Fig. 4.11). The A extract exhibited a homologous series of *n*-alkanes and some branched alkanes, and unexpectedly B extract showed the same. Composition of C extract was similar to the A extract.

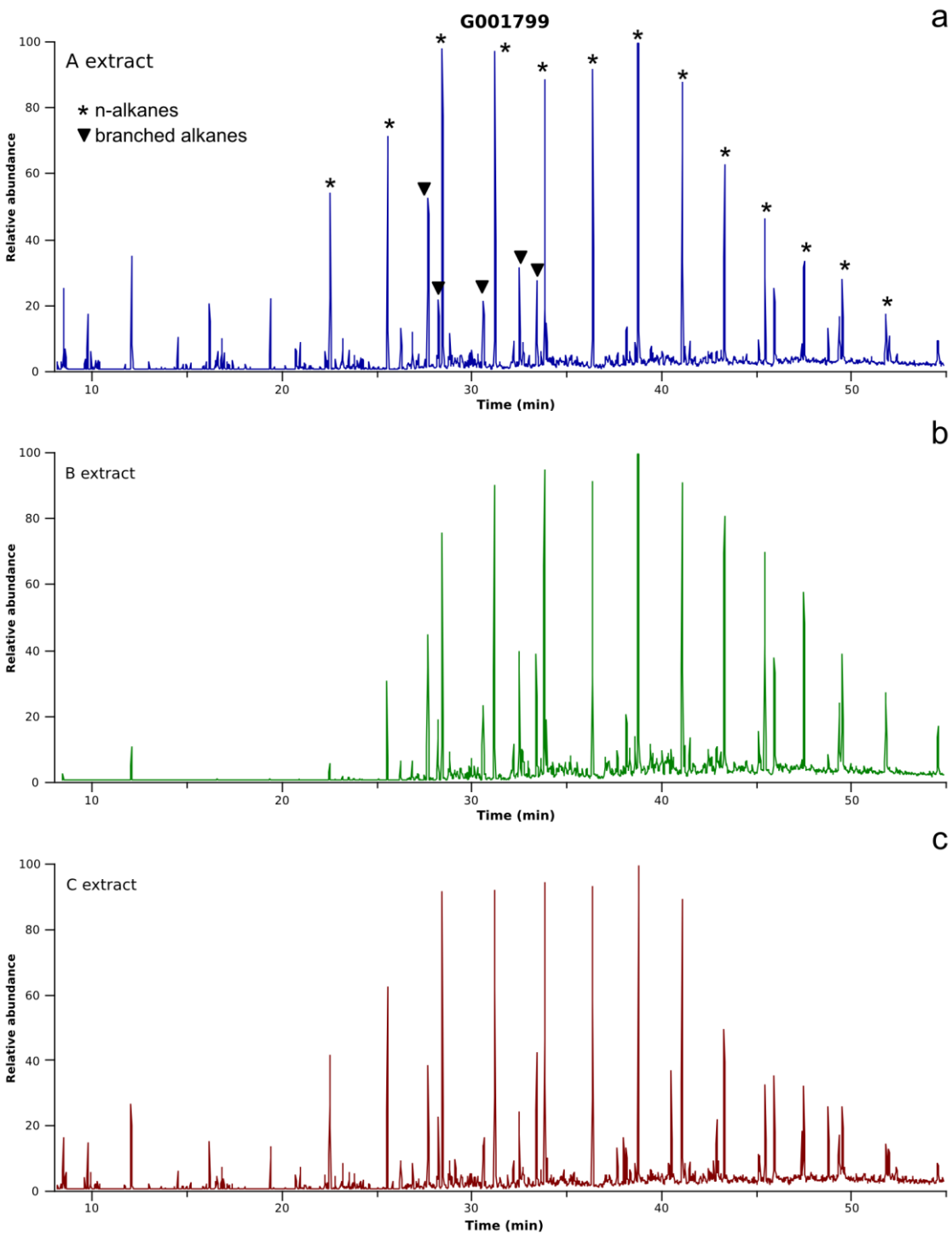


Fig. 4.11 GC-MS chromatograms of A (a), B (b) and C (c) DCM extracts of G001799 sample. These chromatograms were created using m/z 57, 71 and 85 ions, characteristic mass traces.

The amount of hydrocarbons present in B extract was expected to be low as the first LLE step should remove such compounds. The presence of alkanes in B extract may be explained due to emulsions produced during LLE. Emulsion is a suspension of

(ultra)microscopic droplets of one hydrophobic solvent into water, and forms when water extracts and DCM were mixed as part of liquid-liquid extraction. The most common methods to destroy an emulsion are waiting time and addition of salt water, however the latter was discarded to prevent any alteration of the sample. During this first approach, an emulsion was produced and it could have been possible that the waiting time to separate the phases would have been insufficient.

To obtain C extract, LLE was applied to the aqueous sample which had been directly acidified and its composition was similar to A extract, this showed that acidification applied at the beginning of LLE was not effective to extract carboxylic acids only. The same result, similar composition of A and C extracts and presence of alkanes in the B extract was also shown for the other 3 samples (G007923, G004548 and G007022).

Concerning the identification of LMWOA, some saturated carboxylic acids containing 5 to 12 carbon atoms were found in all DCM extracts though they were at trace level compared to the abundance of the alkanes, as exemplified for extracts of G007923 (Fig. 4.12). The homologous series of acids found in the three extracts of G007923 included pentanoic, hexanoic, heptanoic, octanoic, nonanoic, decanoic and dodecanoic acid. Identification was made by the characteristic mass fragmentation pattern (*m/z* values of 60, 73 and 87).

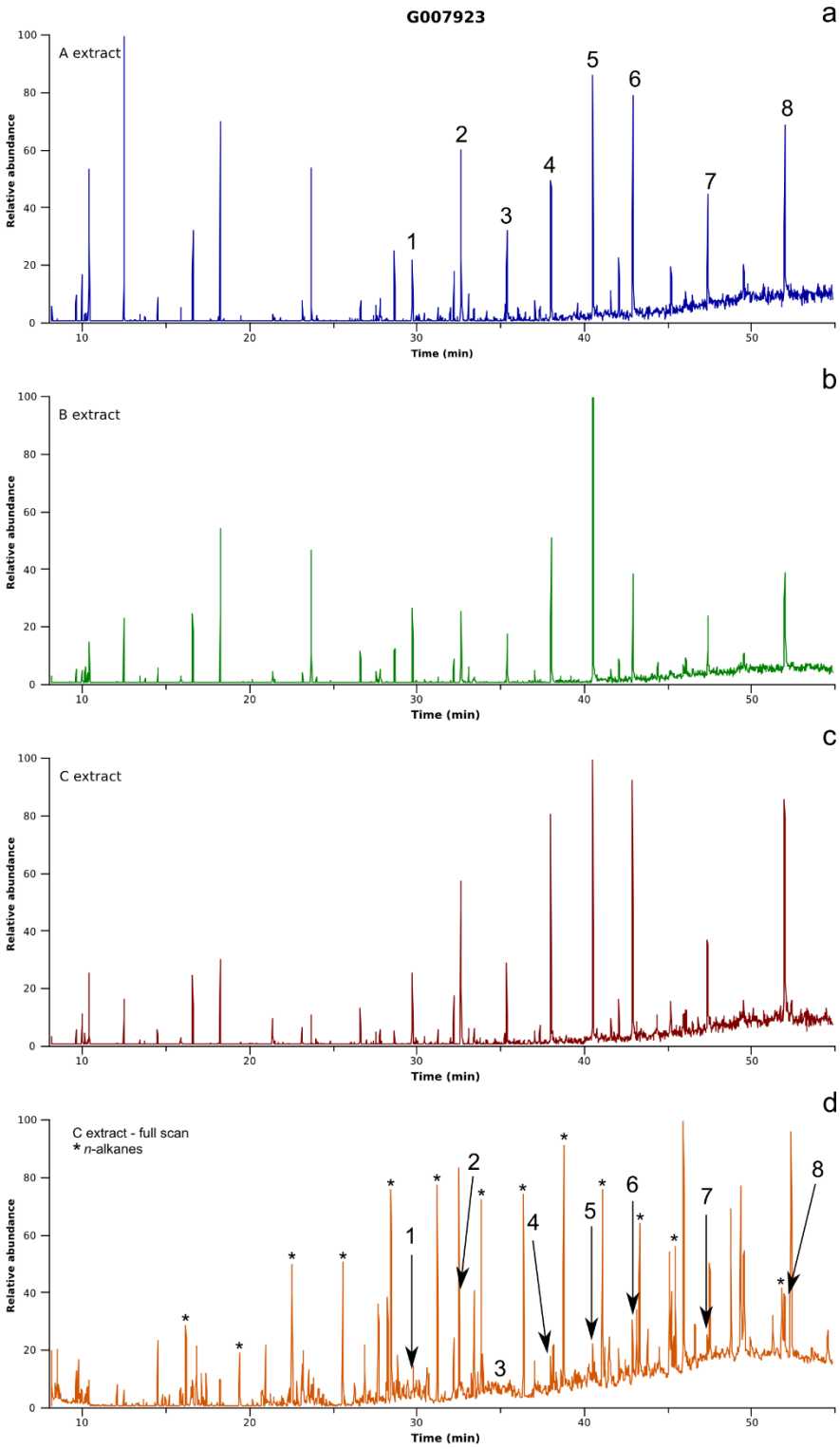


Fig. 4.12 GC-MS chromatograms of A (a), B (b) and C (c) DCM extracts of G007923 sample. These chromatograms were created using *m/z* 60, 73 and 87, except for the one at the bottom. A full scan chromatogram of C extract (d) was also included as a reference to show that abundance of the alkanes was much higher than the one for acids. 1, pentanoic acid; 2, hexanoic acid; 3, heptanoic acid; 4, octanoic acid; 5, nonanoic acid; 6, decanoic acid; 7, dodecanoic acid; 8, tetradecanoic acid.

Fatty acids were expected to be present in the B and C extracts due to the acidification of the water extract that favored the extraction with DCM by decreasing the solubility of the acids in water. The presence of these acids in the A extract can be explained by a simplified equilibrium equation of the acid dissociation (Eq. 4.3) and dissociation constant (pKa) (Eq. 4.5).

$$HA \leftrightarrow H^+ + A^- \qquad \text{Eq. 4.3}$$

$$K_a = \frac{[H^+][A^-]}{[HA]} \qquad \text{Eq. 4.4}$$

$$pKa = pH + \log[HA] - \log[A^-] \qquad \text{Eq. 4.5}$$

When the pH of a solution equals the pKa of the acid, the acid is 50 % dissociated, but Eq. 4.5 shows that regardless of the pH of the solution, a portion of the protonated acid will remain in the solution, even in small quantities. In this way, a suitable extraction method must succeed in effectively concentrating and recovering the analyte in a manner that the protonated portion stays below a certain detection limit. For example, by consecutive extractions using fresh solvent.

The results of this first approach showed that the originally applied LLE procedure needs to be modified to remove the non-polar compounds and improve the extraction of the carboxylic acids.

4.2.2 Second approach

A new procedure was designed by including basification of the sample and applied to the samples with high TOC, namely G007923 and G004548. In this modified procedure, the pH of soxhlet extract was first increased up to 10, followed by three times LLE with DCM. Secondly, the pH of the extracted aqueous solution was lowered to less than 1 and subsequently extracted with DCM three times (Dias *et al*., 2002). This procedure also included a mechanical shaking during LLE (Fig. 4.13) enhancing the DCM-water contact, thus to improve the extraction.

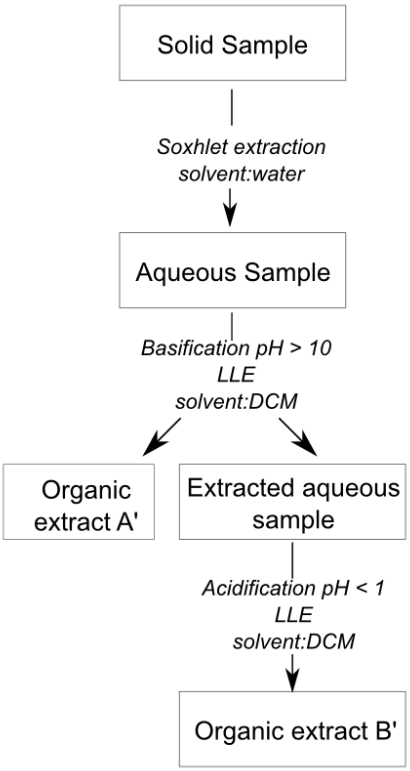


Fig. 4.13 Modified liquid-liquid extraction procedure including mechanical shaking.

This procedure provided a significant reduction of alkanes in the second extract. Fig. 4.14 shows the difference in alkane content between B (first LLE approach, Fig. 4.14a) and B' (second LLE approach, Fig. 4.14b) extracts for sample G004548. While in B extract the alkanes (A extract, gray line) are more abundant than organic acids (red line), the B' extract (A' extract, green line), as desired, shows a clear reduction of alkanes due to the new LLE procedure. It is obvious that the compounds eluting at short retention times, presumably alkanes and some alcohols of low molecular weight are no longer present. A tentative identification of these compounds was made using mass spectra library.

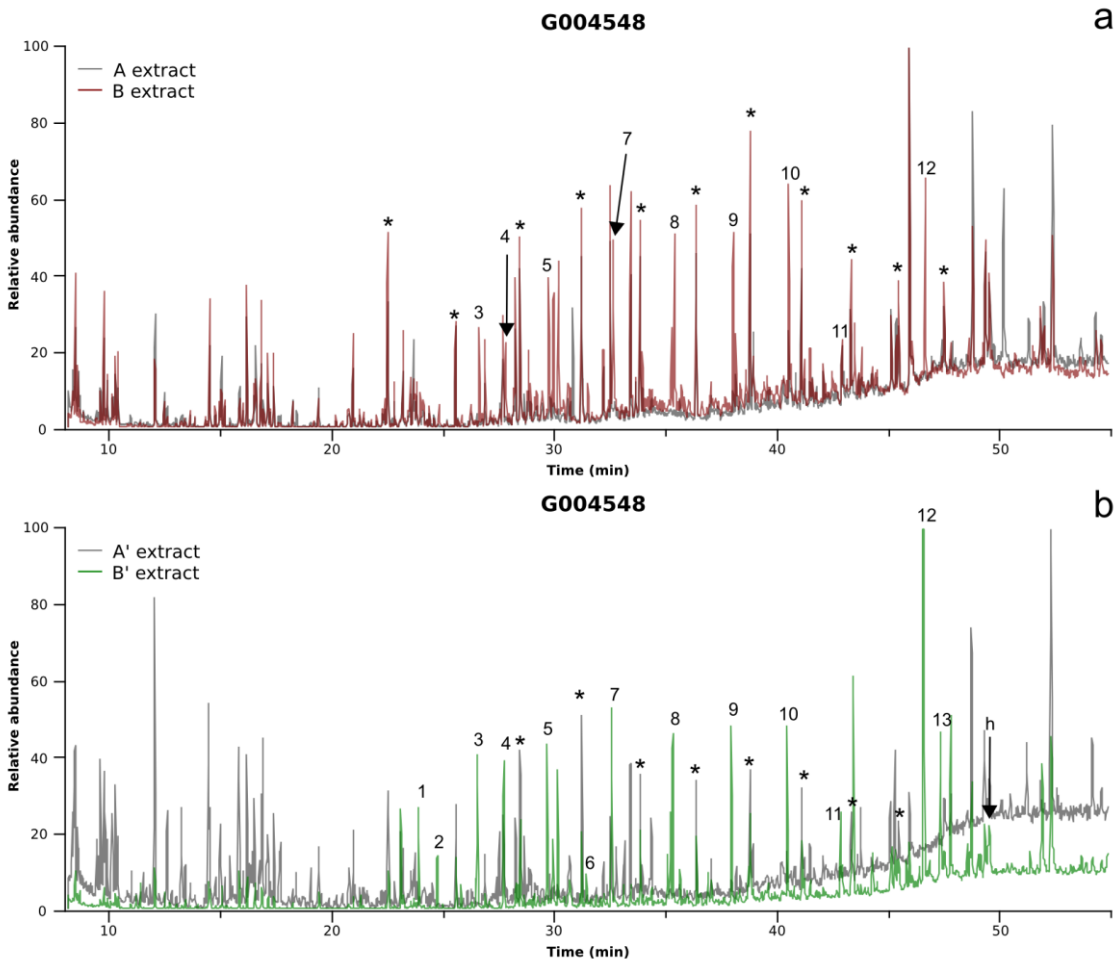


Fig. 4.14 GC-MS chromatograms showing the differences in alkane and LMWOA content in the B (a, first LLE approach) and B' (b, second approach) extracts of G004548. 1, propanoic acid; 2, 2-methylpropanoic acid; 3, butanoic acid; 4, 2-methylbutanoic acid & 3-methylbutanoic acid; 5, pentanoic acid; 6, 4-methylpentanoic acid; 7, hexanoic acid; 8, heptanoic acid; 9, octanoic acid; 10, nonanoic acid; 11, decanoic acid; 12, benzoic acid; 13, dodecanoic acid; h, vanillin; *, alkanes.

Furthermore, this procedure allowed the identification of higher number of carboxylic acids. In addition to the homologous series of *n*-alkanoic acids found previously, propanoic acid, 2-methylpropanoic acid, butanoic acid, 2-methylbutanoic acid & 3-methylbutanoic acid, 4-methylpentanoic acid, dodecanoic acid and vanillin, a phenolic aldehyde precursor of terrestrial organic matter were also identified.

The composition of acids in the extract of G004548 (Fig. 4.14) was similar to the one for G007923 (Fig. 4.15), as both samples showed a homologous series of *n*-alkanoic acids starting from C3 (propanoic acid) up to C12 (dodecanoic acid), some methyl-

branched alkanoic acids and benzoic acid. However the higher relative abundance of vanillin in the extract of G07923 than that in the extract of G004548 may be related to maturity, R_0 of 0.39 and 0.29 respectively (a detailed discussion about composition of water extracts is provided in Chapter 5).

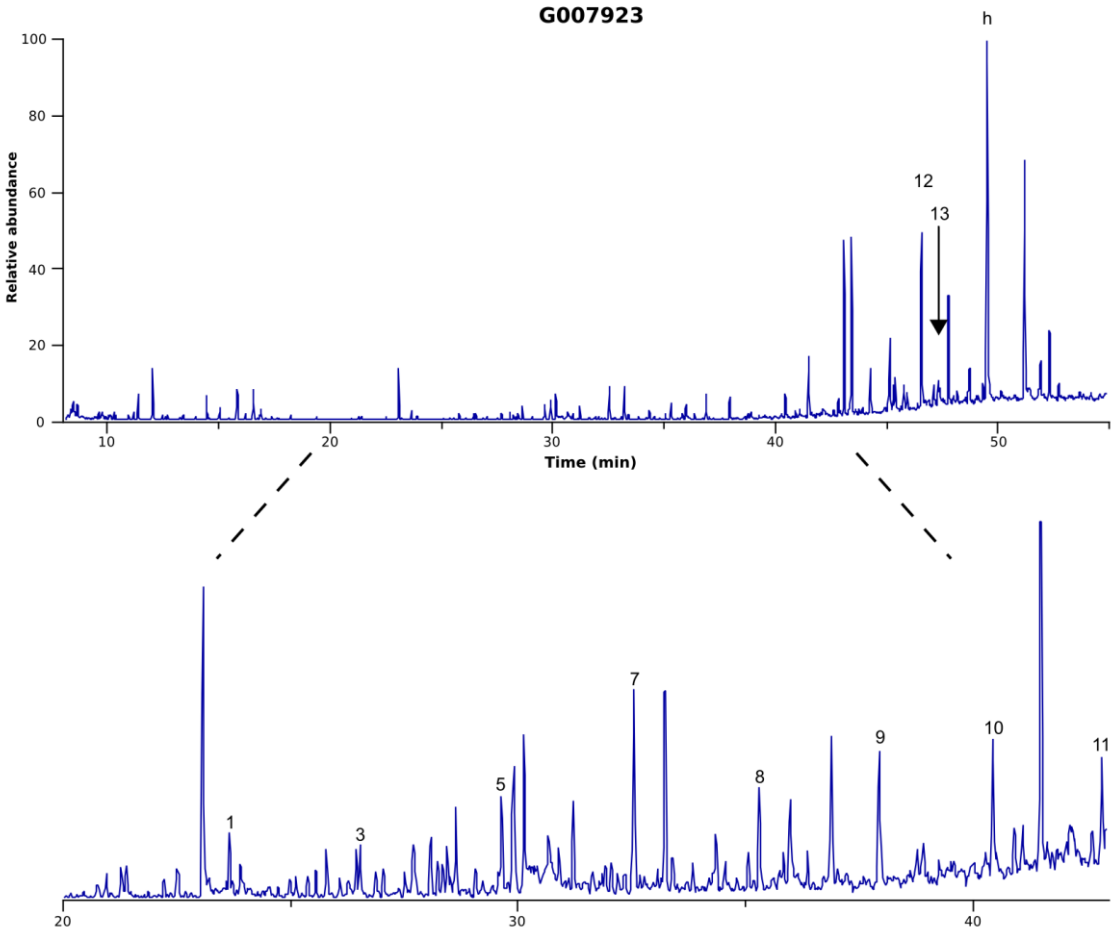


Fig. 4.15 GC-MS chromatograms showing the diversity of LMWOA detected in the B' extract of G007923. See Fig. 4.13 for nomenclature.

With this second approach, a good separation between non-polar hydrocarbons and carboxylic acids was achieved. In the next step the efficiency of the modified LLE procedure was tested, i.e. recovery assessment.

4.2.3 Recovery assessment

The efficiency of the LLE procedure was assessed comparing the amount of analyte before and after extraction using aqueous synthetic samples containing four acids: propanoic acid, hexanoic acid, nonanoic acid and cycloheptanecarboxylic acid. The concentration of acids in stock solution ranged between ca. 0.01 mg/l to 1.8 mg/l depending on their solubility. Five different dilution ratios were tested, namely 10:1, 30:1, 100:1 300:1 and 1000:1, and two shaking times (15 and 60 minutes) were also evaluated. GC-MS was employed for quantification and 2,2-dimethylpentanoic acid served as an internal standard (IS). As previously mentioned, due to difference in sensitivity amongst analytes and the internal standard, a relative response factor (RRF) was calculated for each compound (Fig. 4.9).

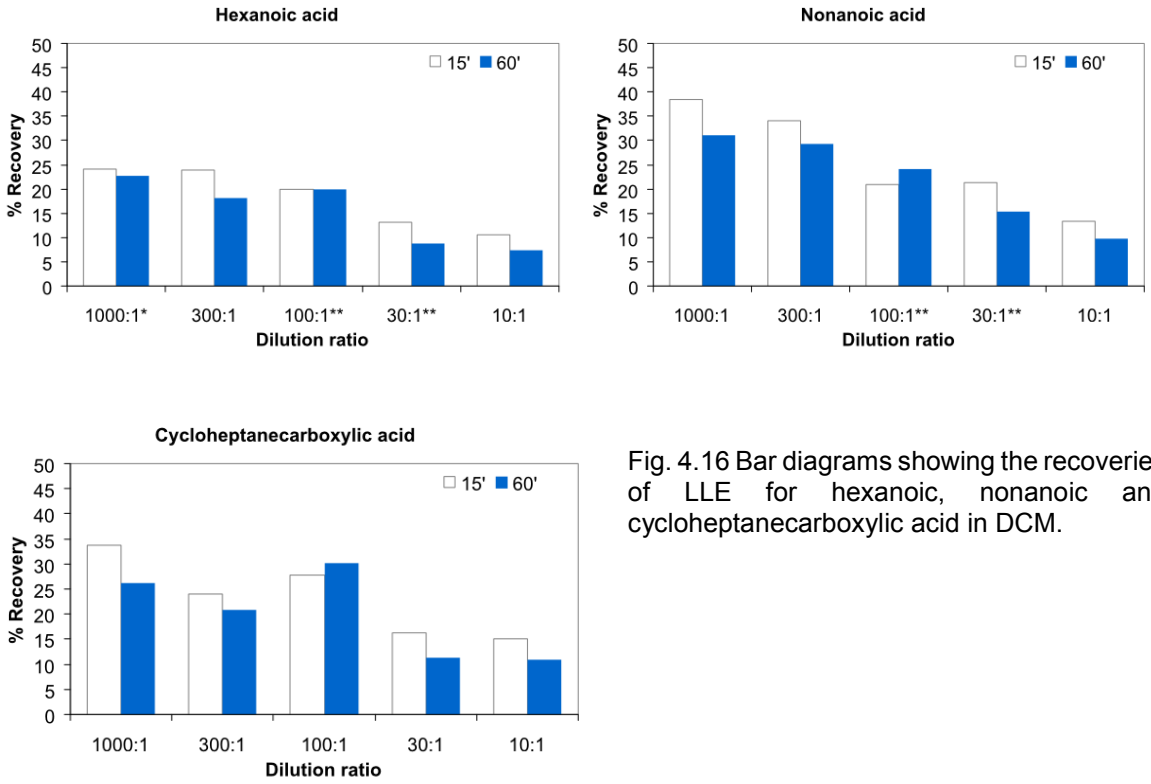


Fig. 4.16 Bar diagrams showing the recoveries of LLE for hexanoic, nonanoic and cycloheptanecarboxylic acid in DCM.

In general, a 15-min shaking gave slightly higher recoveries than the 60-min shaking. Propanoic acid, quantified from dilution ratio 100:1 to 10:1, had the lowest LLE recovery (<1%). For all acids, the recovery was in inverse proportion to concentration, which means that high recoveries were present for low concentrations. For a 15-min shaking time, the recovery was between 10.7-24.2% for hexanoic acid, between 13.3-38.4% for nonanoic acid, while cycloheptanecarboxylic acid had values between 15.1-33.8%. (Fig. 4.16).

The low recovery was related to i) DCM was not appropriate for extraction of CA evaporation during shaking; ii) sorption on the sodium sulphate bed during drying. To solve this drawback the LLE procedure was modified by employing diethyl ether (DEE) as solvent, the amount of sodium sulphate was drastically reduced; mechanical shaking was also eliminated. For the new test, the aqueous standards contained two compounds (propanoic and hexanoic acid), with concentrations ranging between 0.05-1mg/l. The internal standard was also substituted using 1-methylcyclohexanecarboxylic acid, as 2,2-dimethylpentanoic acid exhibited co-elution with 3-methylpentanoic acid, an acid found in the water extracts of coals.

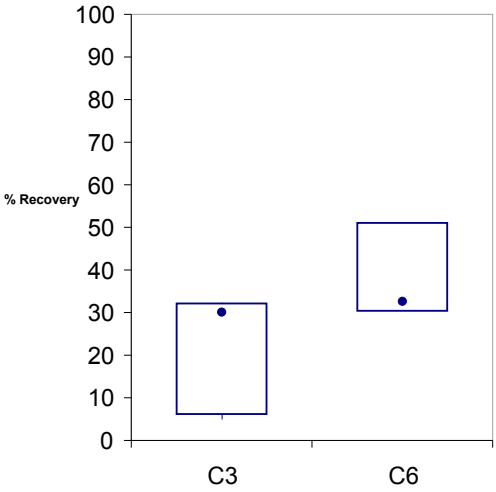


Fig. 4.17 Boxplots that display the recovery of LLE using DEE as solvent.

High recoveries are important because they proportionally affect the sensitivity. Moreover, a low recovery increases the detection limit (Kou and Mitra 2003). From a practical point of view, in many cases sample amount could also be limited. Thus high recoveries may provide good quantification basis with small amount of sample. Fig.

4.17 shows the recovery for propanoic acid was between 10-30% and for hexanoic acid it was between 30-50%. Recovery for both acids was improved in comparison with the one obtained using DCM as solvent, but it still was low, so a new technique needed to be employed.

4.3 Solid-phase extraction

The selected solid-phase extraction (SPE) procedure was proposed by Jurado-Sánchez *et al.*, (2010). During the establishment of the SPE procedure, several problems needed to be solved: a) the setting of the SPE apparatus; b) degradation of the extracted compounds in methanol; c) selection of a new eluent; d) selection of internal standard; and, e) the appearance of a hump in the GC-FID chromatograms when using the new eluent. The following subsections describe the conditions and strategies employed to optimize the SPE protocol.

4.3.1 Setting up the SPE apparatus

The semiautomatic SPE device was built with the following materials: a syringe pump and a standard HPLC column packed with the proposed sorbents and glass wool (Fig. 4.18). The procedure was as follows: a) injection of acidified 50 ml aqueous sample, b) drying of column with 50 ml air, c) elution of analytes with 200 µl methanol and 30 ml air, the organic extract was collected in an amber glass vial (2 ml) with Teflon-lined screw cap; and d) rinsing the column with 1 ml of Millipore water and 10 ml air. The syringe pump supplied flow rate of 3 ml/min. The methanol extract would later be analyzed by GC-MS and/or GC-FID for quantification.

For calibration aqueous standards were used in concentrations ranging between 1.25 µg/l and 160 µg/l. Such solutions contained 11 compounds: propanoic, 2-methylbutanoic, pentanoic, hexanoic, 2-methylhexanoic, cyclopentylacetic, nonanoic, decanoic, benzoic and dodecanoic acid. The extracts were then analyzed with GC-FID and the peak identification on GC-FID chromatograms was made by comparison of the retention times of standards in methanol.

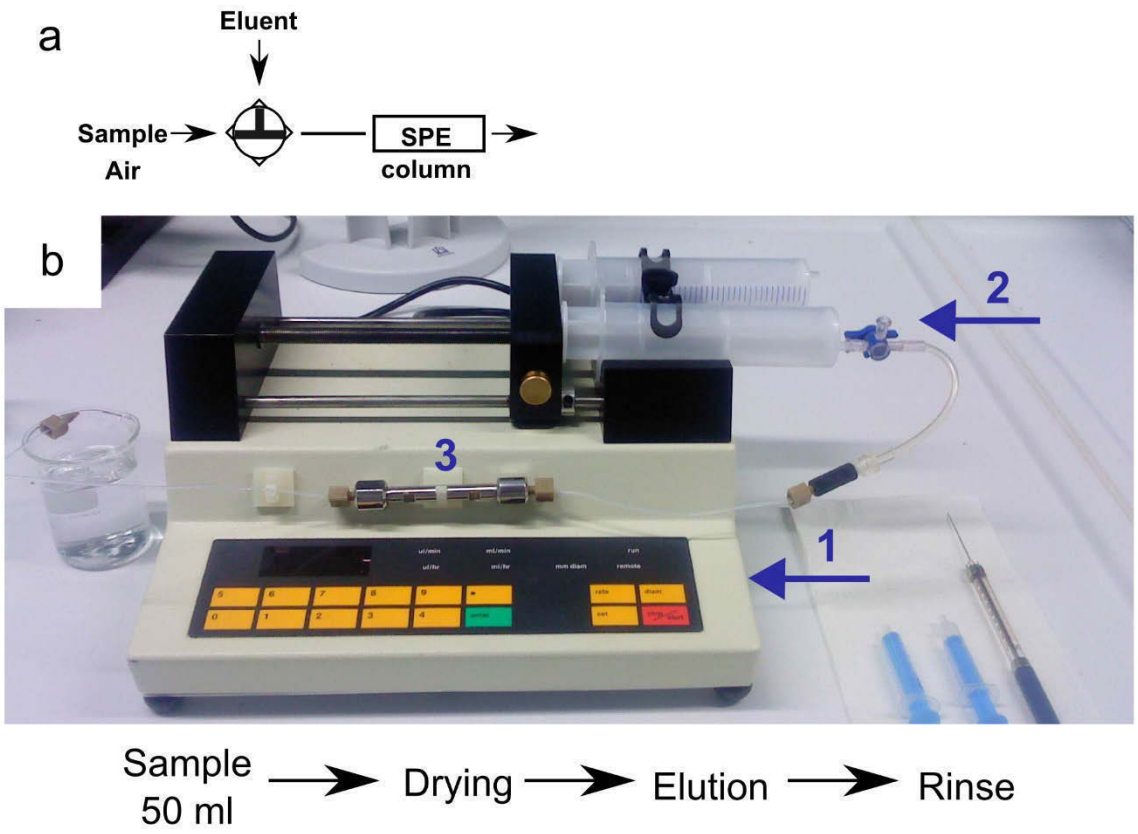


Fig. 4.18 Semiautomatic SPE apparatus. a) Schematic diagram of the SPE apparatus showing the flow of the fluids: water and air are injected through the syringe, and the eluent (methanol) through the valve. b) Picture of the SPE apparatus and its parts: 1) syringe pump; 2) 3-way valve and 3) SPE column.

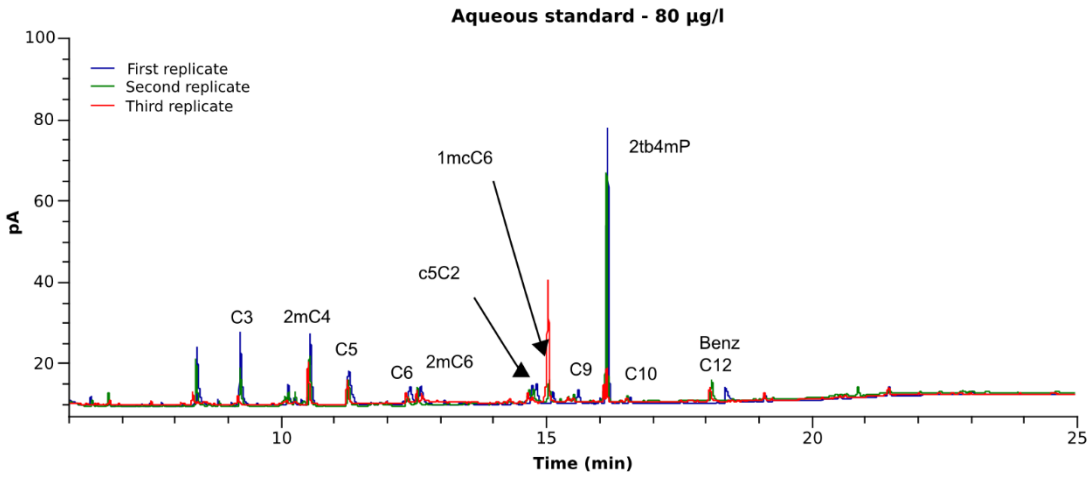


Fig. 4.19 GC-FID chromatograms of methanolic extracts of aqueous standards (80 µg/l). Random variations of peak area and peak shape of individual acids. C3, propanoic acid; 2mC4, 2-methylbutanoic acid; C5, pentanoic acid; C6 hexanoic acid; 2mC6, 2-methylhexanoic acid; c5C2, cyclopentylacetic acid; C9 nonanoic acid; C10, decanoic acid; Benz, benzoic acid and C12, dodecanoic acid.

The SPE calibration curves showed significant variations in peak areas of acids and internal standard amongst SPE replicates (Fig. 4.19), which meant that any calculation based on this SPE system would be unreliable. The reason for such variations was assumed to be possible voids in the filling of the SPE column, thus the addition of inert sand along with the sorbents was tested and a 200 µl-loop valve was installed in the SPE apparatus (Fig 4.20) to optimize the supply of the eluent.

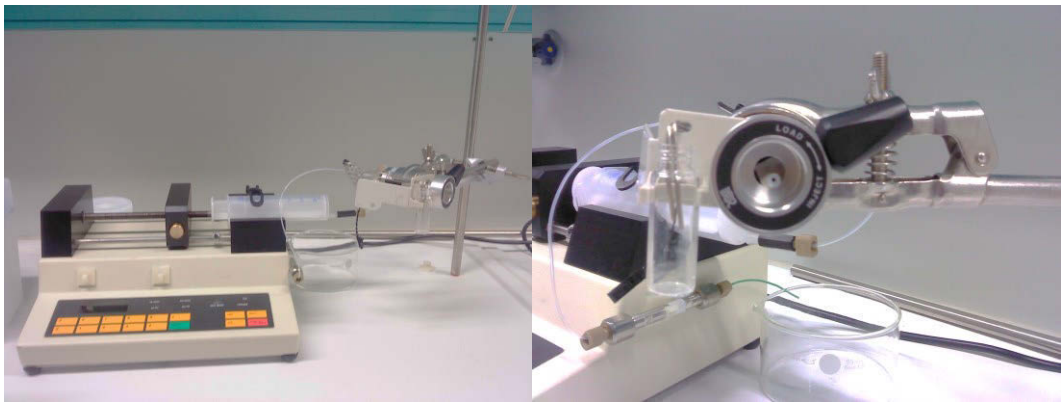


Fig. 4.20 Installation of a 200 µl-loop valve in the SPE apparatus. The new valve was included to reduce the variations in the amount of eluent (methanol).

The effect of the sand, packed along with the sorbents, was evaluated by comparison of the peak area of the internal standard between replicates and blanks (procedure applied to acidified Millipore water) in the calibration curves. For the third calibration curve, the aqueous standards with the highest concentrations (80-160 µg/l) could not pass the column and the flow through the SPE column was no longer possible even when the flow rate for the aqueous solution was reduced to 2 ml/min. This could presumably be due to compaction of the filling (sand and sorbents). Fig. 4.21 shows that the peak areas of internal standard varies slightly in the same run (calibration curve) but the changes are more evident amongst replicates. It can also be seen that the third curve (S3) exhibits the lowest IS peak area. Such variations might be related either to the SPE system or the FID chromatograph. Through replicates of GC-FID analysis of SPE extracts and standards in methanol, GC-FID resolution decay became recurrent and important affecting the reliability of the calculations (Fig. 4.22).

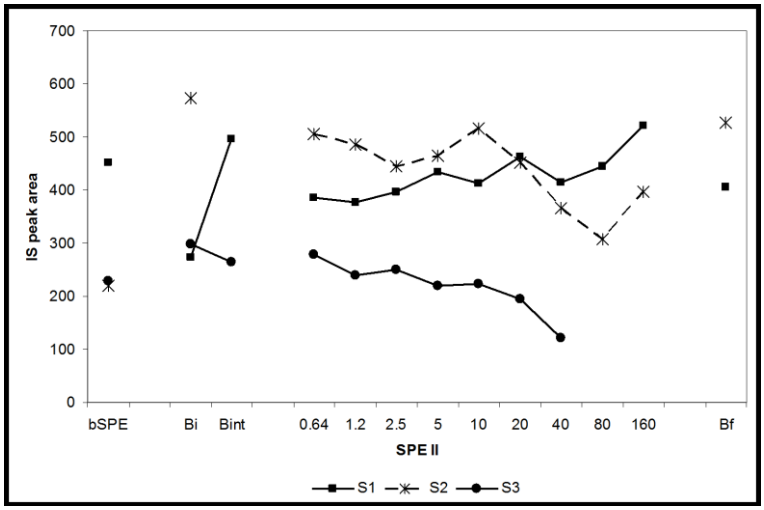


Fig. 4.21 Variations of peak areas of internal standard after adding inert sand to the SPE column filling along with the sorbents ENVI-18 and LiChrolut® EN. SPE procedure was sequentially applied to (i) the aqueous standard solutions from the lowest to the highest concentration, (ii) blank and finally (iii) the standard solutions. S1-S3, first, second and third SPE run; Bi, Bint and Bf stand for initial, intermediate and final blank, respectively.

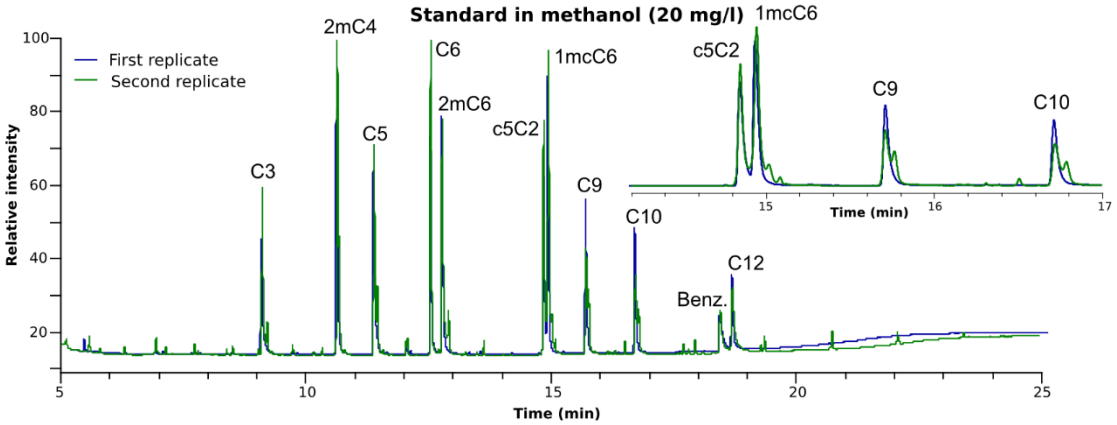


Fig. 4.22 GC-FID resolution decay. In the chromatograms can be seen the variation of the peak shape in two different GC-FID runs of the same methanolic standard indicating the decay of GC column. See Fig. 4.19 for nomenclature.

Every part of the chromatograph and injection system was checked and the chromatographic column was re-installed several times. It was assumed that the INNOWax column was already degraded and should be replaced more often than expected. Thus, just one problem was left to be solved: prevention of the SPE column compaction.

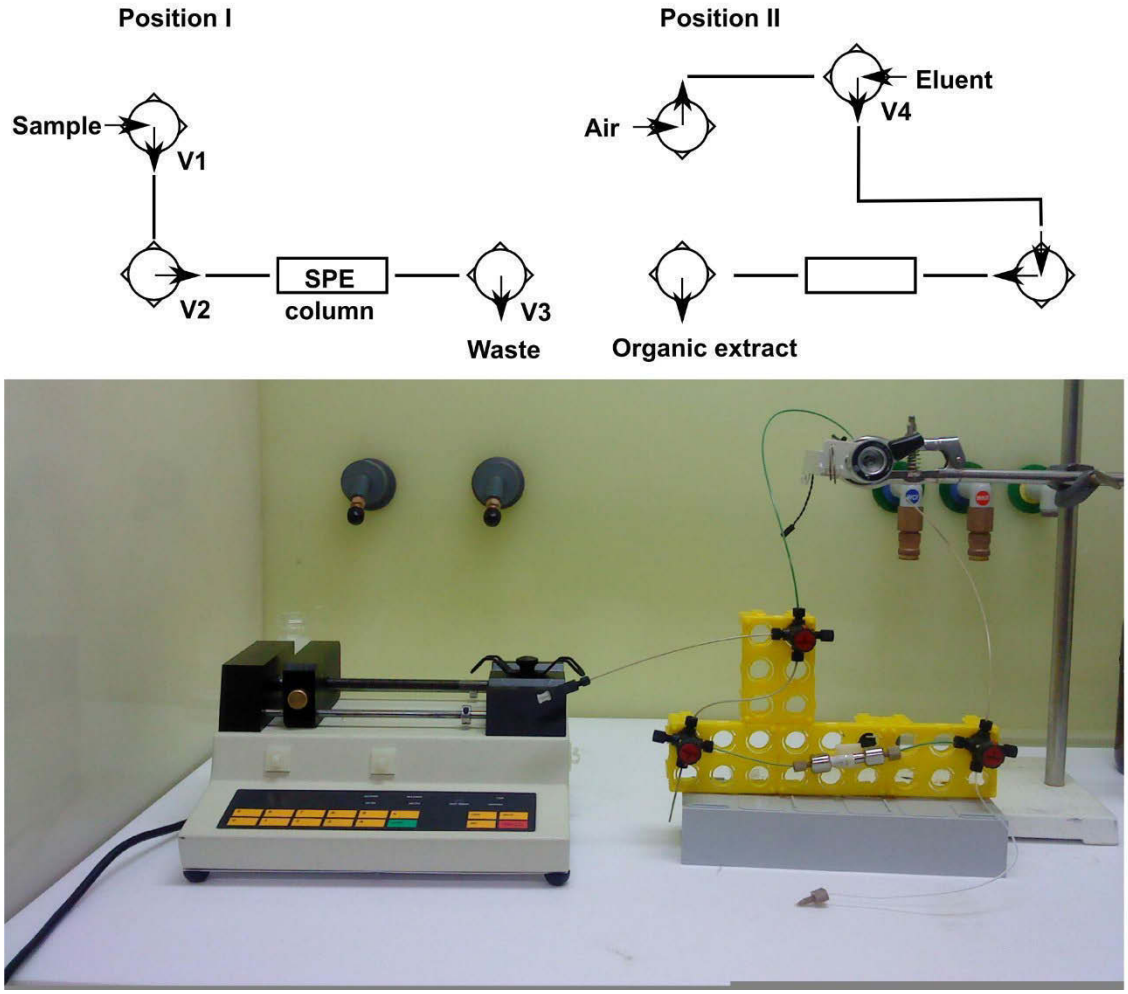


Fig 4.23 New SPE configuration. Three valves were installed to control the flow direction in the SPE column (a). The acidified aqueous sample (pH<2, 5M HCl) flows through the column (Position I) for column drying the flow is inverted by shifting the direction of valves elution (Position II). The eluent is supplied through a 200µl-loop injection valve (V4).

A new design of the SPE system was established (Fig. 4.23) which included the use of three 3-way switching valves (in addition to the one for eluent supply) and a new SPE column with smaller inner diameter, matching the size given by Jurado-Sánchez *et al*., (2010). The new configuration also involved a flow inversion: the sample would flow in a direction opposite to the drying and elution steps. The procedure was as follows: a) loading of 50-ml acidified aqueous sample at flow rate of 2 ml/min (Position I, Fig. 4.23); b) changing of flow direction on the three valves and drying of SPE column with 50 ml-air (flow rate of 3 ml/min) (Position II, Fig. 4.23); and, c) injection of 200 µl of methanol spiked with the internal standard through the supply valve by pumping 30 ml air with a flow rate of 3 ml/min. The organic extract containing the target analytes

was collected in an amber glass 2-ml vial with Teflon-lined septa in screw cap. The syringe pump supplied the chosen flow rates.

With these modifications, the triplicated calibration curves by triplicate no longer exhibited any problem related to column compaction.

4.3.2 Degradation of acids in methanolic extracts

Once the correct functioning of flow and column at SPE apparatus was achieved, a persistent appearance of new peaks was detected in the SPE extracts (Fig. 4.24). Through repetitions of the SPE procedure with blanks and 10-compound aqueous synthetic samples, it was observed that the new peaks were related to degradation of the acids. A feasible reason was that the acidic remains of HCl in the column were not negligible and catalyzed the transformation of the fatty acids into their methyl esters in the methanolic extract (Fig. 4.25)

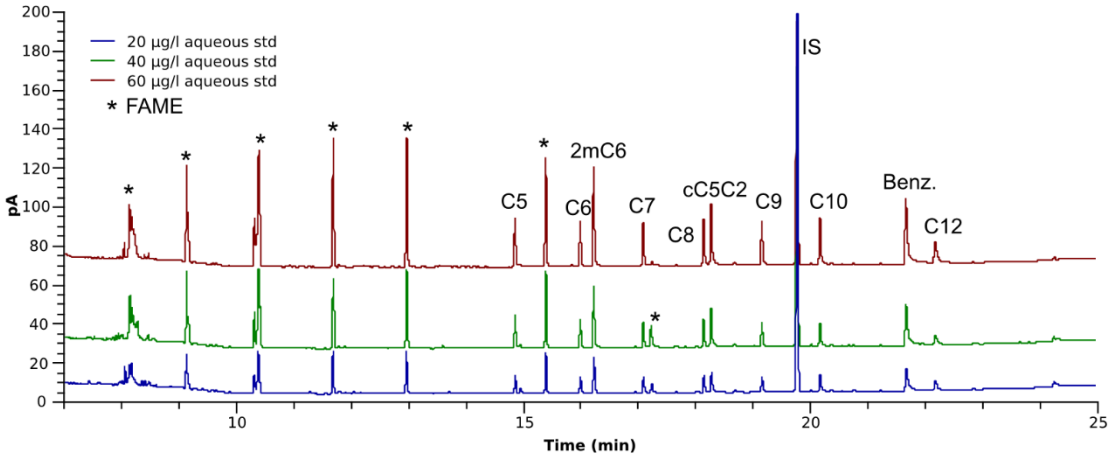


Fig. 4.24 GC-FID chromatogram showing fatty acid methyl esters (FAME's) in the methanolic SPE extracts. See Fig. 4.19 for nomenclature.

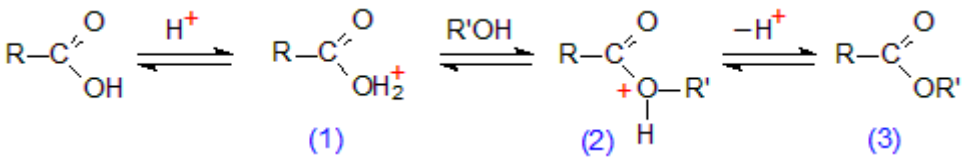


Fig. 4.25 Acid-catalyzed esterification of fatty acids. The reaction starts with the protonation of the acid (1), then the newly produced oxonium ion undergoes an exchange reaction with an alcohol to give an intermediate product (2), which in turn can lose a proton to become an ester (3). In the presence of a large excess of the alcohol, the equilibrium point of the reaction is displaced so that esterification may proceed to completion (Christie, 2011).

To solve the degradation several strategies were tested as follows: a) cleansing of hydrochloric acid remains; b) preservation of methanolic extracts; c) substitution of HCl for acidification of aqueous sample; and, d) substitution of the eluent.

Cleansing of acidic remains

Once the acidified aqueous sample was applied to the SPE column, it was rinsed with millipore water and then the procedure was followed as previously established. For the rinse a volume of 2 ml water was tested but a negative effect on the recovery of the fatty acids was observed. As shown in Fig. 4.26, peak areas of all acids were significantly smaller with rinse. This may have been related to desorption of the fatty acids from the sorbents to the water, therefore this cleansing was not suitable to prevent FAMEs formation.

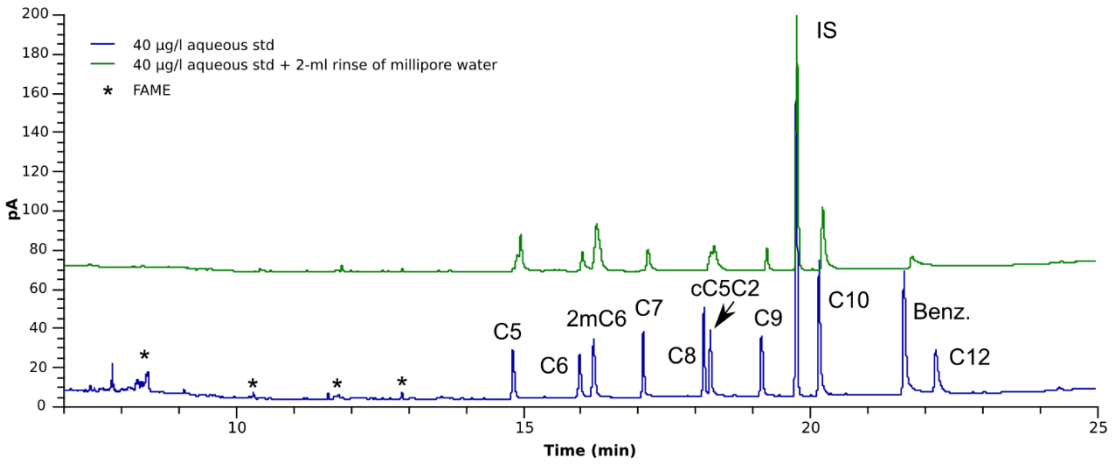


Fig. 4.26 GC-FID chromatogram showing the effect of a 2-ml rinse in the SPE column in the methanolic extract right after the sample was injected. GC-FID analyses were immediately made after the SPE procedure. See Fig. 4.19 for nomenclature.

Preservation of methanolic extracts

Temperature

To avoid the acid-catalyzed esterification of fatty acids, the storage of the methanolic extracts at -25°C was also tested but it showed that the low temperature did not prevent the formation of FAME's either (Fig. 4.27).

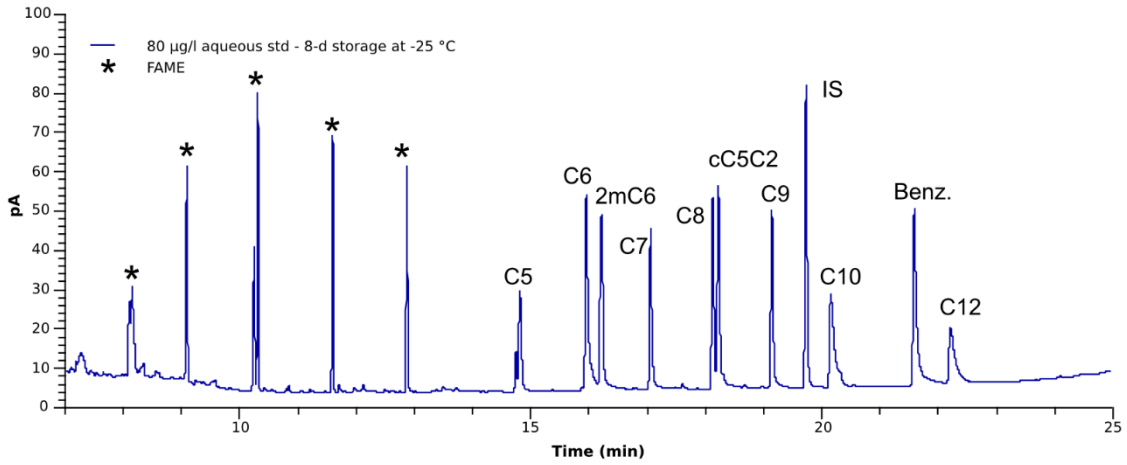


Fig. 4.27 GC-FID chromatogram showing the effect of 8-day storage of methanolic extract at -25°C. See Fig. 4.19 for nomenclature.

Changing pH

As a strategy to stop the formation of FAME's, conditioning of the pH of methanolic extracts was also tested. In order to neutralize the HCl remains that acidified the methanol, triethylamine (TEA) a common base used in organic synthesis, was added to the organic extracts. GC-FID analyses were carried out on methanolic extracts after an 8-day storage at two different temperatures, room temperature (RT) and 4°C. It was seen that TEA marginally prevented the formation of methyl esters but also significantly reduced the quality of the chromatograms (e.g. peak splitting), most likely because the methanolic extracts became too basified by the amine (Fig. 4.28 dark red and green lines) leading to sorption of carboxyl groups onto the INNOWax column. Then a re-acidification of the organic extract was tested to improve the resolution of the GC-FID analyses. This was done using benzenesulfonic acid or hydrochloric acid (35%). Though the peak splitting was slightly prevented (from 18 min, red line), both acids produced a hump in the middle of the chromatogram what would complicate the quantification of acids; in the case of HCl, the hump completely masked the fatty acid peaks (Fig. 4.28 olive line). Thus, this strategy was unsuccessful in preventing the degradation of the organic extracts.

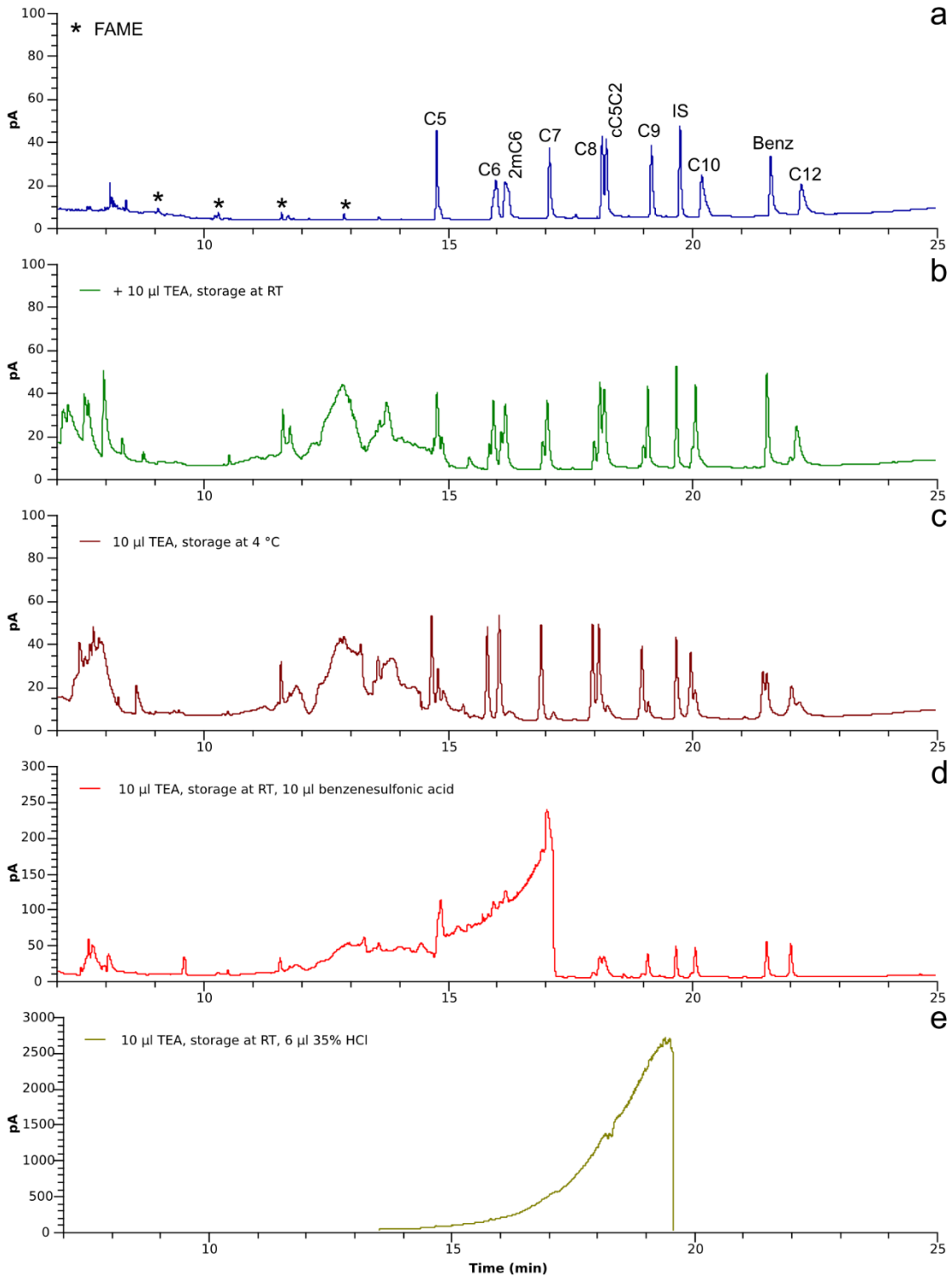


Fig. 4.28 GC-FID chromatograms showing the effect of conditioning the pH of methanolic extracts (80 µg/l aqueous sample). GC-FID run made after first hours of the SPE procedure. It is included as reference (a). Chromatogram of the methanolic extract with 10 µl of TEA stored for a few days at room temperature (RT) (b) and at 4°C (c). Chromatogram of the methanolic extract with 10 µl of TEA, stored at room temperature for a few days with addition of benzenesulfonic acid (d) or hydrochloric acid (e) before GC-FID analysis. See Fig. 4.19 for labelling.

Substitution of HCl to acidify the aqueous sample

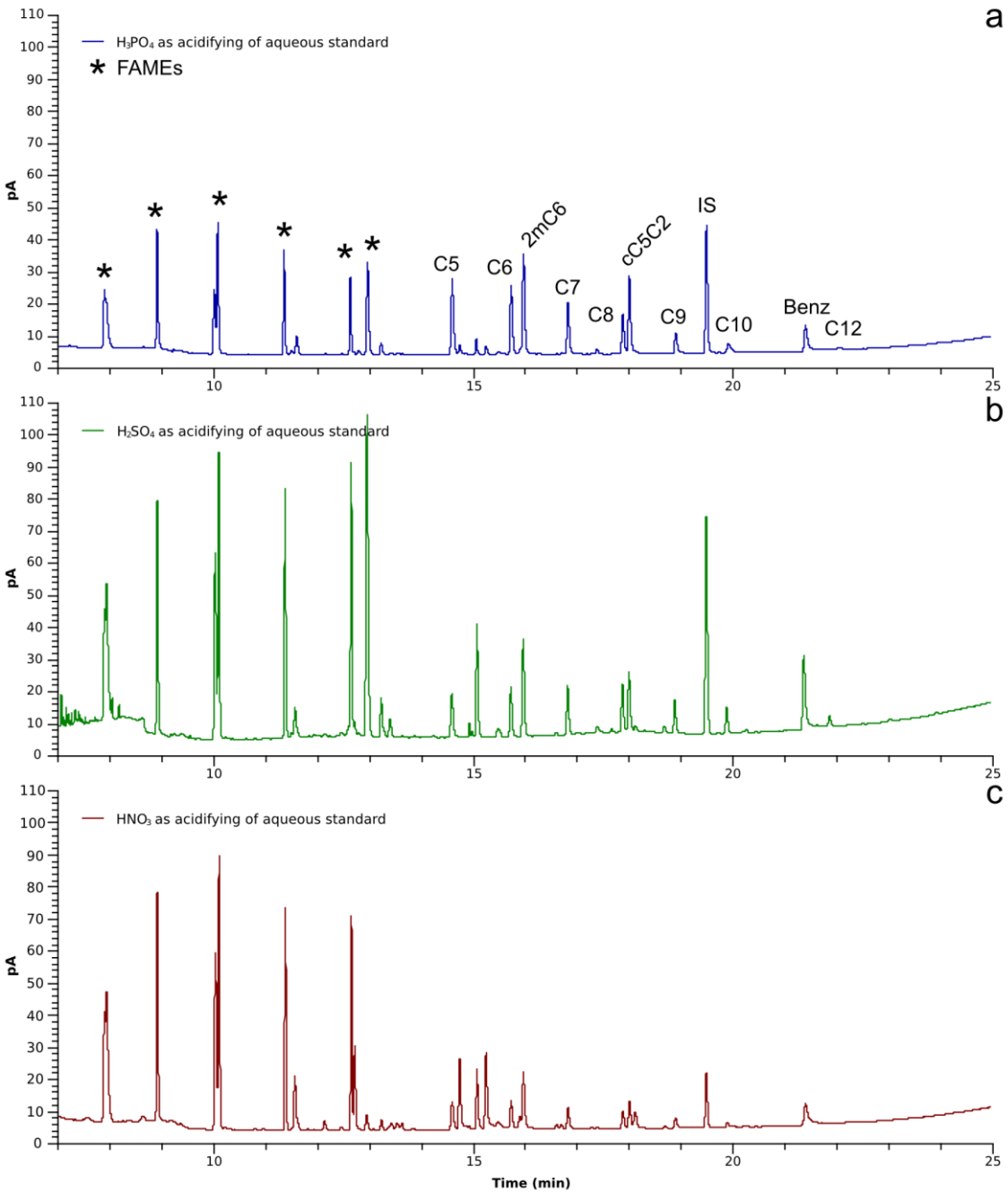


Fig. 4.29 GC-FID chromatograms showing the effect of acid employed to acidify the aqueous sample. The acids tested were phosphoric acid (a), sulfuric acid (b) and nitric acid (c). See Fig. 4.19 for labeling.

The acidification of the aqueous sample allows the fatty acids to remain in their protoned form optimizing their sorption onto the chosen SPE reversed phases, namely

Supelclean ENVI-18 and LiChrolut® EN. As another approach to solve the problem described above, phosphoric, sulfuric and nitric acid were used as substitutes of hydrochloric acid for the acidification of aqueous sample. This was the only change in the SPE procedure (Fig. 4.23). But as shown in Fig. 4.29, the acid remains in the column still favored the esterification of the fatty acids and FAMEs were produced even after substitution of HCl. Differences in the intensities of the fatty acids between chromatograms were thought to be related to how the three tested acids may interact with the chromatographic column, an INNOWax column based on polyethylene glycol (PEG). The more acidic the eluent, the less adsorption of acids occurs in the chromatographic column. The acid strength increases as $HNO_3 < H_3PO_4 < H_2SO_4$, and it is seen that the chromatogram of the HNO_3 test shows the lowest peak areas and peak shapes of the target compounds.

4.3.3 Substitution of eluent

The selection of a new eluent was the most important modification to the SPE procedure after the observation that HCl remains in the column resulted in an undesirable acid-catalyzed esterification of fatty acids in methanol extracts. Several organic solvents were tested as eluent: ethanol, 2-propanol, acetone, methyl formate, ethyl acetate, 1,4-dioxan, tetrahydrofuran, acetonitrile (pure), acidified acetonitrile (10% acetic acid), dimethyl sulfoxide and dimethylformamide. For the test an aqueous standard was employed containing 10 acids with a concentration of ca. 80 µg/l and acidified with 5 M HCl. After SPE, the organic extracts were stored at 4°C. GC-FID analyses were made right after SPE with a waiting time between 24-48 h.

The GC-FID and GC-MS chromatograms of the SPE extract are the result of the process as a whole: SPE procedure and GC-FID (or GC-MS) performances. If most procedural conditions were kept the same (e.g. sample volume) it can be assumed that the SPE efficiency will only depend on the elution strength of the solvent. For reversed-phases the eluotropic strength increases as solvent polarity[1] decreases (Wells 2003). Supelclean ENVI-18 and LiChrolut EN are suitable for moderately polar and very polar

[1] Solvent polarity refers “the capacity of a solvent for solvating dissolved charged or neutral, apolar or dipolar species.” (Reichardt, 2011)

compounds respectively (Sigma-Aldrich Co. 1998; Merck-Millipore 2015), thus alcohols (hydrophilic solvents) are common eluents for these sorbents.

Several aspects were considered for the selection of the new eluent. The three most significant criteria were a) formation of esters or any other by-product after 24 h or later; b) peak solvent overlapping the fatty acid peaks; and, c) good signal response for high boilers, i.e. from nonanoic acid. Chromatographic resolution, peak shape and signal intensity within the whole chromatogram were also taken into account.

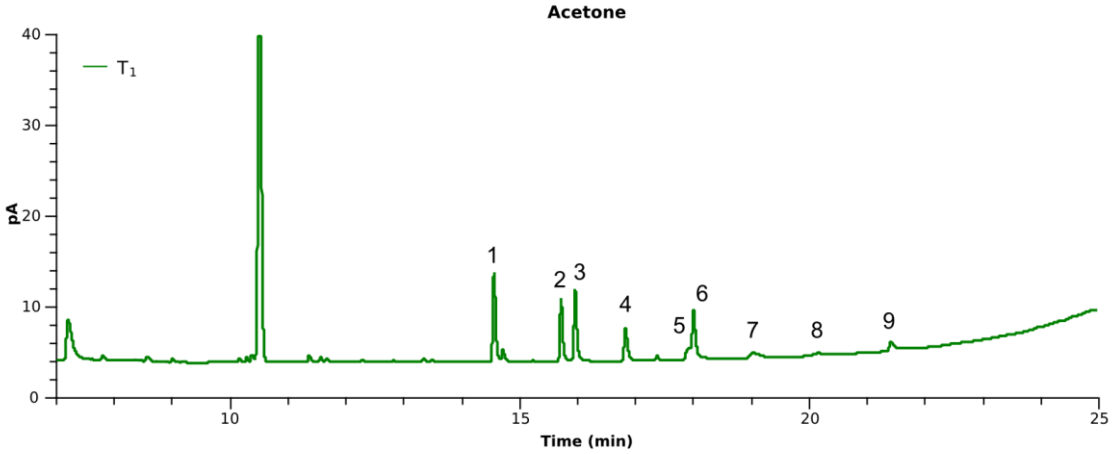
Table 4.3 Qualitative assessment of GC-FID chromatograms for tested eluents

Eluent	Additional peaks	Low resolution[a]/ overlap of fatty acid peaks[b]	Signal[c] intensity	E^N_T[d] (polarity)
Methanol	yes		medium	0.762
Ethanol	yes		medium	0.652
2-propanol	yes		medium	0.546
Acetone	yes	low resolution	low	0.355
Methyl formate	yes		medium	0.346
Ethyl acetate	yes		low	0.228
1,4-dioxan	yes		medium	0.164
Tetrahydrofuran	yes	low resolution	medium	0.207
Acetonitrile (pure)	yes	low resolution	low-medium	0.460
Acidified acetonitrile (10% acetic acid)	yes	low resolution	medium	
Dimethyl sulfoxide	yes	overlap	very high	0.444
Dimethyl-formamide	no		very high	0.386

GC-FID analysis carried out on an INNOWax column. [a]Especially between octanoic and cyclopentylacetic acid peaks, and/or low response of fatty acids from nonanoic acid, [b]peak solvent masking the ones of the fatty acids, [c]compared with methanol. [d]Normalized $E_T(30)$ solvent polarity parameter (Reichardt and Welton 2010) For explanation see text.

Presence of more than ten peaks corresponding to the carboxylic acids contained in the aqueous standard was considered as an indicator of by-products. The new eluent should not produce any of them, its mere presence was relevant and no further identification was performed. Regarding response, the intensity of fatty acids shown with methanol (the one purposed by Jurado-Sánchez, *et al.* 2010) as eluent was used as reference to evaluate this parameter. Due to the nearly similar retention times of

octanoic and cyclopentylacetic acid, attention was also paid to the good chromatographic resolution of both compounds.



$$CH_3OCH_3 + Cl_2 \rightleftharpoons CH_3COCl + HCl$$

$$CH_3COHCl + Cl_2 \rightleftharpoons ClCH_2COCH_2Cl + HCl$$

Fig. 4.30 GC-FID chromatogram SPE extract in acetone (eluent). The additional peaks could be a product of the chlorination of acetone during GC-FID analysis. 1, pentanoic acid; 2 hexanoic acid; 3, 2-methylhexanoic acid; 4, heptanoic acid; 5, octanoic acid; 6, cyclopentylacetic acid; 7, nonanoic acid; 8, decanoic acid; 9, benzoic acid and 10, dodecanoic acid. T1 refers to GC analysis made 24 h after SPE procedure.

Table 4.3 summarizes the evaluation of the different solvents tested as eluent in SPE procedure. To some extent, most of the solvents favored the formation of by-products, seen as additional peaks in the chromatograms, except for dimethylformamide. The additional peaks observed in ethanol are likely esters, the formation of which is due to residual HCl. The chromatogram of the isopropanol eluent also exhibited unidentified peaks, but they were smaller when compared to the ones seen using methanol or ethanol as eluent. Acetone exhibited low recovery of the acids and additional peaks (Fig. 4.30). Although the largest additional peak did not overlap any of the analytes, it may interfere when working with natural samples. It is known that acetone may produce chloroacetone or dichloroacetone in the presence of gaseous chloride if appropriate temperature and acetone/chlorine mole ratios exist (Markoš et al. 1999). These conditions may exist during GC-FID analysis and thus, the observed additional peak could possibly be related to that reaction.

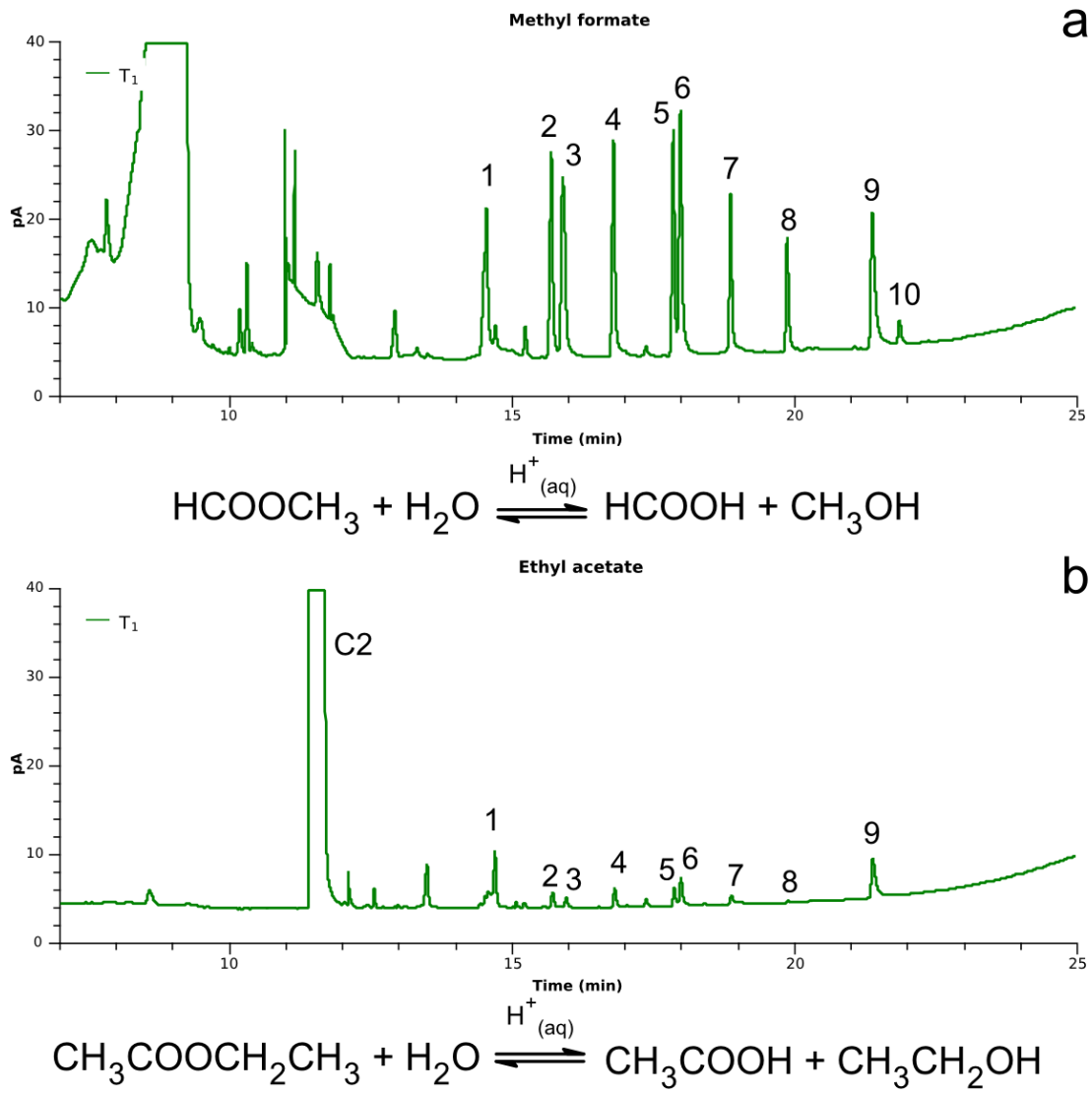


Fig. 4.31 GC-FID chromatograms of methyl formate and ethyl acetate tested as eluent in SPE. The remains of HCl could have catalyzed the hydrolysis of the esters, obtaining as by-products formic acid and methanol from methyl formate, and ethanol and acetic acid from ethyl acetate. Acetic acid peak was identified by its retention time. Numbers from 1 to 9 represent the target analytes. See Fig. 4.30 for nomenclature.

When using esters for elution, the additional peaks could be the products of acid-catalyzed hydrolysis, formic acid and methanol derived from methyl formate, and acetic acid and ethanol from ethyl acetate (Fig. 4.31); in turn the alcohols may also have undergone acid-catalyzed esterification to some extent. In these chromatograms, an apparently significant difference between recoveries can also be seen. The higher signal response of methyl formate may be related to the fact (which rarely occurs) that this solvent desorbed the fatty acids left in the SPE column by acetone (eluent previously tested which presented a low recovery). Two factors may explain the results

obtained by ethyl acetate as eluent: water remains in the SPE column and its immiscibility in water (Mitra, 2003). The aqueous sorbed layer in the SPE phases could hinder the ethyl acetate wetting the sorbents and thus desorption of the analytes.

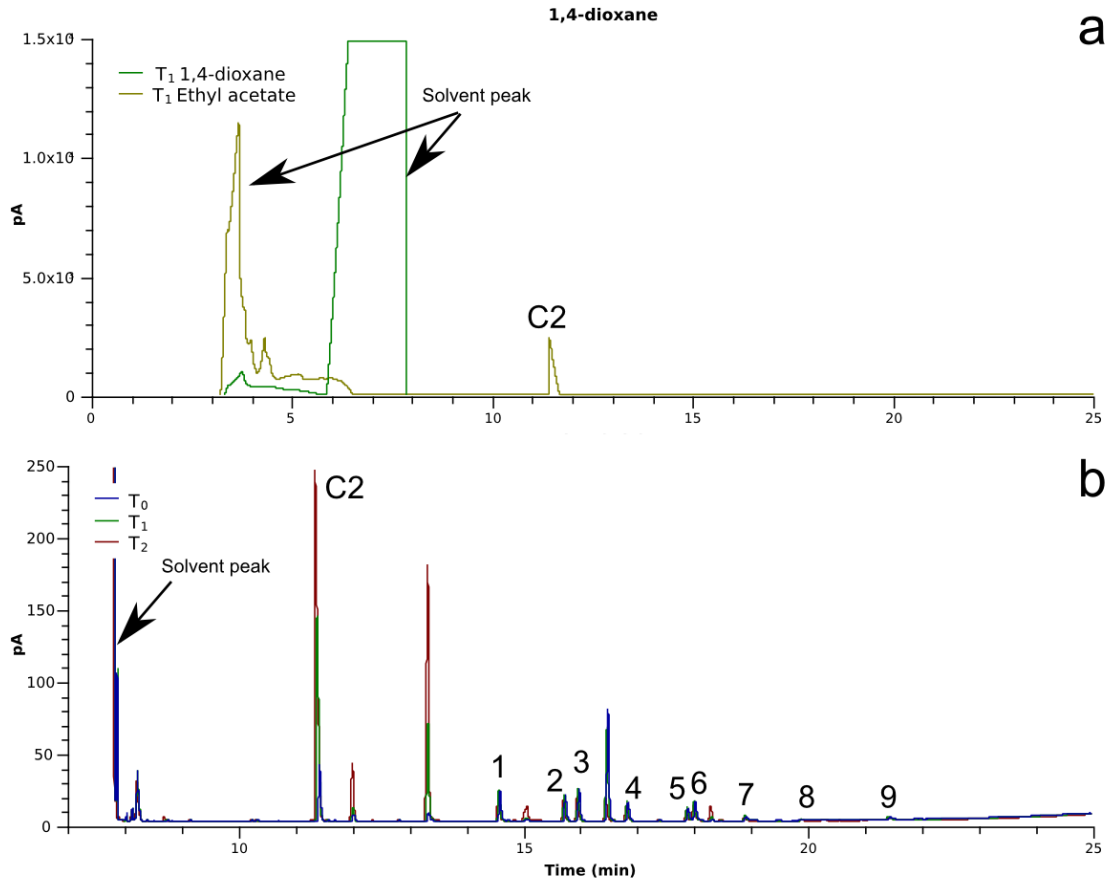
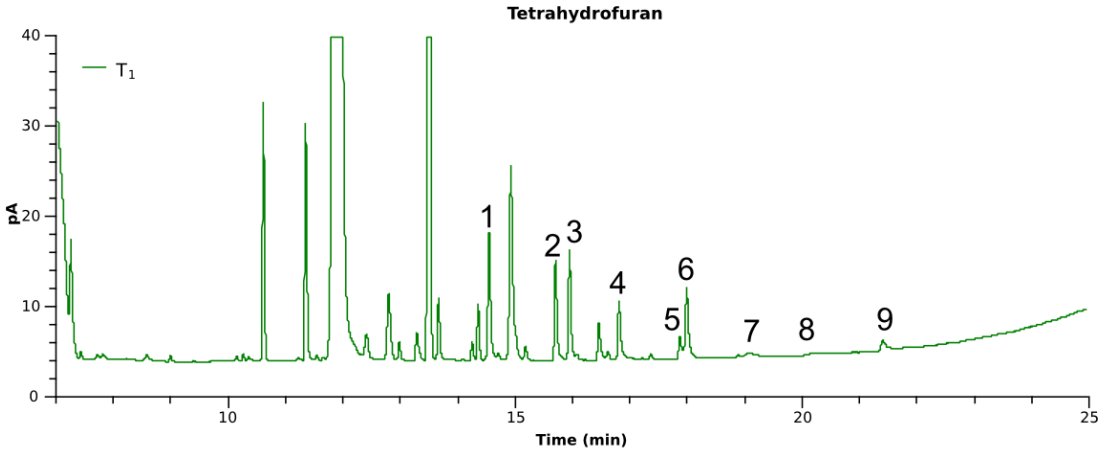


Fig. 4.32 GC-FID chromatograms of 1,4-dioxane as eluent in SPE. Ti represents GC-FID analyses made at different times: T0 right after SPE procedure, and T1 and T2, at 24 and 72 h respectively. This sequence shows the likely increasing formation of by-products. Acetic acid peak was identified by the retention time. Numbers from 1 to 9 represent the target analytes. For nomenclature, see Fig. 4.30.

In case of 1,4-dioxan as eluent, chromatogram showed additional peaks, which may be due to SPE column abnormally retaining a portion of ethyl acetate (Fig. 4.32a) and with the ester remains undergoing hydrolysis during storage (Fig. 4.32b). If the amount of the ethanol was sufficient, there could have been an additional acid-catalyzed esterification reaction as well.



$$(CH_2)_4O + HCl \longrightarrow Cl(CH_2)_4OH \xrightarrow[-H_2O]{HCl} Cl(CH_2)_4Cl$$

Fig. 4.33 GC-FID chromatogram of tetrahydrofuran as eluent in SPE. For nomenclature see Fig. 4.30.

Tetrahydrofuran SPE extract also exhibited additional peaks (Fig. 4.33). Some of them may be related to the acidic cleavage of tetrahydrofuran by hydrochloric acid remains. The first reaction would be opening of tetrahydrofuran ring to give 4-chlorobutane followed by the formation of 1,4-dichlorobutane (Becker, 1977).

In the case of acetonitrile, hydrochloric acid could have hydrolyzed the organic nitrile to acetic acid (Fig. 4.34). The slightly higher recovery exhibited by acidified acetonitrile (10 % v/v acetic acid) would be due to the fact that acidification improves the chromatographic signal response of organic acids.

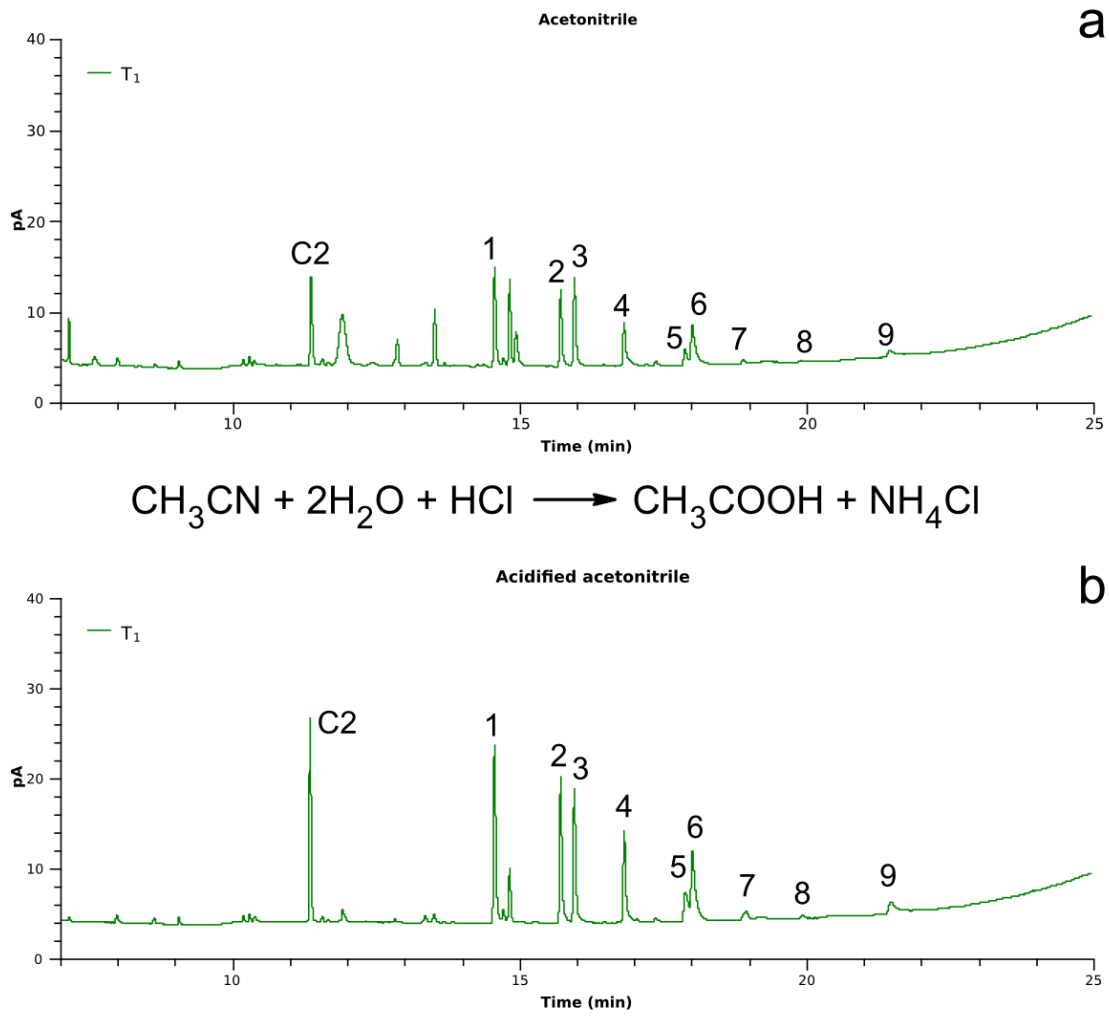
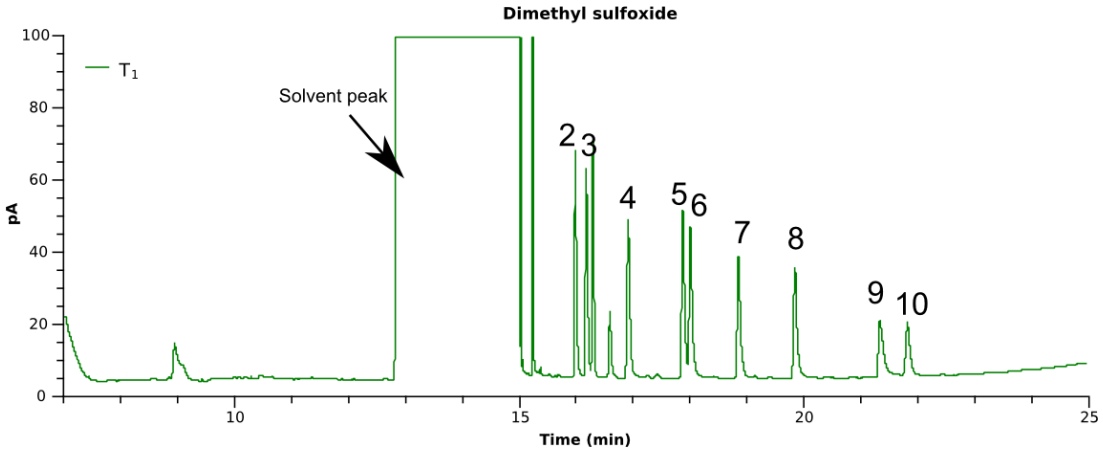


Fig. 4.34 GC-FID chromatograms of acetonitrile (a) and acidified acetonitrile (b) as eluents in SPE. In this case hydrochloric acid could have catalyzed the hydrolysis of acetonitrile into acetic acid. For nomenclature see Fig. 4.30.

The SPE extract with dimethyl sulfoxide (DMSO) as eluent showed additional peaks as well (Fig. 4.35). The thermal degradation of this solvent to form paraformaldehyde, dimethyl sulfide and under certain conditions, dimethyl disulfide has been reported. HCl is assumed to have low influence on the decomposition (Gaylord 2005). Another fact shown by this chromatogram is that the DMSO peak overlaps the one of the pentanoic acid, which would mean the masking of one target analyte by the SPE method.



$$H_3CSOCH_3 \longrightarrow H_3CSH + HCHO \longrightarrow (HCHO)_x$$

$$2H_3CSH + HCHO \longrightarrow (H_3CS)_2CH_2 + H_2O$$

$$2H_3CSH + CH_3SOCH_3 \longrightarrow H_3CSSCH_3 + H_3CSCH_3 + H_2O$$

$$2H_3CSOCH_3 \longrightarrow H_3CSO_2CH_3 + H_2CSCH_3$$

Fig. 4.35 GC-FID chromatograms of dimethyl sulfoxide as eluent in SPE. For nomenclature see Fig. 4.30.

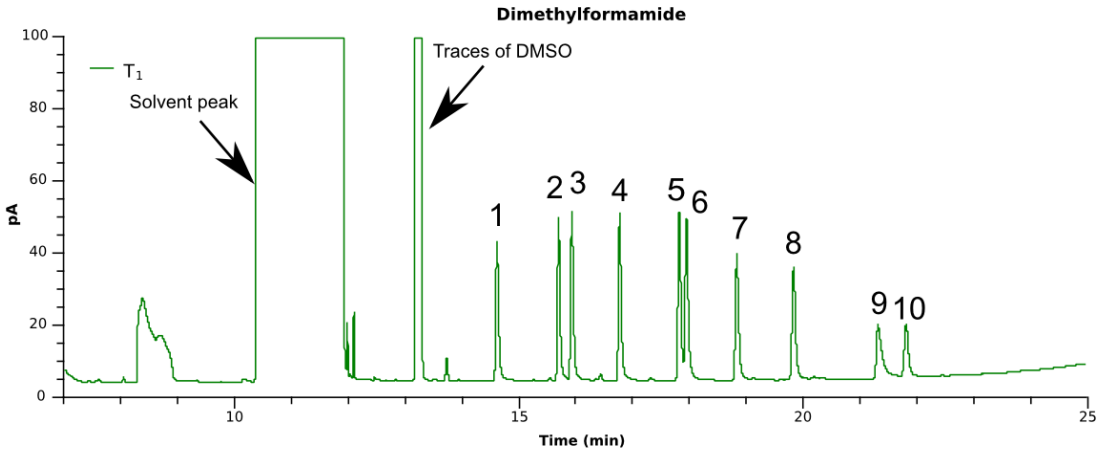


Fig. 4.36 GC-FID chromatogram of dimethylformamide as eluent in SPE. For nomenclature see Fig. 4.30.

Amongst all solvents, dimethylformamide (DMF) was chosen as new eluent, because there were no additional peaks in the chromatogram and the solvent peak did not overlap the ones of fatty acids. Moreover, the signal response of high boilers was very good, the peaks of nonanoic, decanoic, benzoic and dodecanoic acid were large

enough for further quantification (Fig. 4.36). After some additional tests with standard aqueous solutions and in natural samples, it was decided to use acidified DMF (10 % v/v acetic acid) as eluent to improve the chromatographic response by acidification.

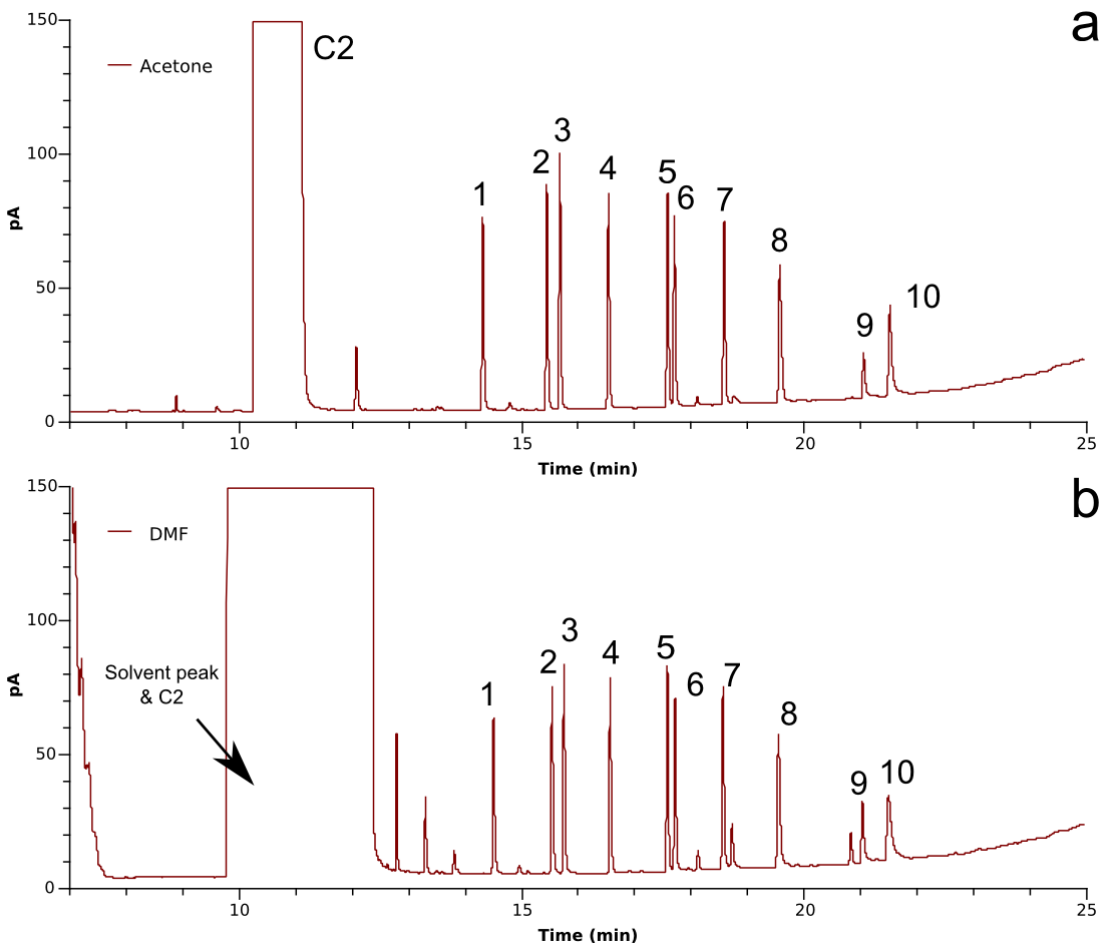


Fig. 4.37 GC-FID chromatograms of standards (20 mg/l) in acetone (a) and dimethylformamide (b). For nomenclature see Fig. 4.30.

Based on the chromatograms a question arises: why do acetone, acetonitrile, DMSO and DMF exhibit significant differences in their recoveries when these solvents are of similar polarity (Table 4.3)? Further, they all are reported to be miscible with water in all proportions.

GC-FID analyses of standards containing the 10 carboxylic acids in acidified solvents showed that the signal was similar amongst standards at same concentration, proving

that the differences observed during eluent test were due to the SPE procedure (Fig. 4.37).

Reversed-phase SPE sorption works on the basis of van der Waals forces (Wells 2003), thus, the observed differences in the recovery are related to the strength of the interactions between sorbent, analyte and eluent. There are two types of van der Waals forces: dispersion forces and dipole-dipole interactions. Dispersion forces refer to the attractions between molecules due to temporary fluctuating dipoles and the dipoles induced onto other molecules. Dipole-dipole interactions are due to the permanent dipole built by the difference in electronegativity amongst the elements in a molecule (Clark, 2012). The boiling point may be a good indicator of the strength of dispersion forces: the stronger the forces, the higher the boiling point (Clark, 2014). The molecular shape also affects these forces: the temporary dipole is stronger in larger molecules.

Table 4.4 Boiling point, solvent strength of the tested eluents and solubility of benzoic acid.

Solvent	Boiling point (°C)[a]	Solvent strength (Al_2O_3)[b]	Measured solubility of benzoic acid (M)[c]
Water	100	Very high	
Methanol	64.5	0.95	2.867
Ethanol	78.3	0.88	2.634
2-propanol	82,4	0.82	2.381
Dimethyl sulfoxide	189	0.75	5.873
Acetonitrile	81.6	0.65	0.858
Ethyl acetate	77	0,58	1.99
Tetrahydrofuran	65-66	0.57	3.367
Acetone	56.2	0.56	2.355
1,4-dioxan	101.5	0.56	3.088
Methyl formate	31.5		
Dimethyl-formamide	153		5.287

[a]Data from MSDS by Merck Millipore at www.merckmillipore.com. [b]Reichardt and Welton (2010) [c]Bradley *et al.* (2015).

Amongst the solvents tested, methyl formate has the lowest boiling point (31.5 °C) and dimethyl sulfoxide the highest (189 °C), while acetone and methanol hold similar values, 56.2 °C and 64.5 °C respectively (Table 4.4). Thus, it can be assumed that the

interactions between the carboxylic acids and DMSO or DMF are much stronger than the ones between acetone or with tetrahydrofuran (65 – 66 °C), providing an explanation for the higher recovery for high boilers. Additionally, the strength of such interactions can be related to solubility as well, the highest solubility of benzoic acid occurs in DMSO and the lowest in acetonitrile (Table 4.4). Likewise, the solubility of linear organic acids in methanol decreases with an increase in the carbon chain length (Table 4.5, Bradley *et al*., 2010).

Given the diverse interactions taking place in the column between the sorbents, solvents (water and eluent) and analytes, the solvent strength[2] ($\varepsilon°$) provides another potential reason for differences in recovery. Amongst the solvents tested, DMSO is more hydrophilic than acetone and 1,4-dioxane, the least hydrophilic compounds, hydrophilicity may contribute to a more effective recovery of the analytes. With this in mind, the significantly smaller peaks for nonanoic, benzoic and decanoic acid in most solvents (e. g. Figs. 4.30 and 4.34) could be also taken as an indicator of low sorption (poorly recovered) because they are more hydrophobic than the target acids with shorter carbon chain. Hydrophobicity in linear carboxylic acids increases with an increase in the length of the carbon chain.

Table 4.5 Solubility of some carboxylic acids in methanol

Linear acids	Solubility in methanol (M)
Propanoic acid	13.41
Butanoic acid	10.94
Pentanoic acid	9.19
Hexanoic acid	7.98
Octanoic acid	6.31
Undecanoic acid	4.0

Source: Data from Bradley *et al.,* (2010).

The data show that the heaviest fatty acids of the series are also affected during GC-FID. As shown in Fig. 4.36, benzoic and dodecanoic acid peaks are smaller than the

[2] The solvent strength indicates the free energy of adsorption of the solvent per unit area of adsorbent with unit activity, it is defined $\varepsilon° = 0$ for n-pentane on alumina (Reichardt, 2010).

ones of nonanoic and decanoic acid. As explained above (Section 4.1), the low signal for high boiler fatty acids (i.e. from nonanoic acid) is related to the GC-FID conditions. To prevent the deterioration of the column, the maximum temperature of the GC column should be set to 250 °C, very close to the boiling points of nonanoic and benzoic acid and lower than the boiling point of decanoic, dodecanoic acid. This condition favors the phenomenon known as discrimination (McMaster, 2007a; McMaster, 2007b).

4.3.4 Chromatographic problems with DMF as eluent

After DMF was chosen as eluent, the calibration curves of SPE system were built and the procedure was applied to aqueous extracts of natural samples. The method was also tested by means of solutions of unknown concentration of some LMWOA. It was during this time that an anomalous hump appeared in the GC-FID chromatograms (Fig. 4.38). The hump was first thought to be contamination in the SPE column because it had not been observed in several previous tests; it just showed up in the chromatogram of the last SPE run (Test Solution 2) and in its GC-FID replicate. With this in mind, the SPE column was refilled.

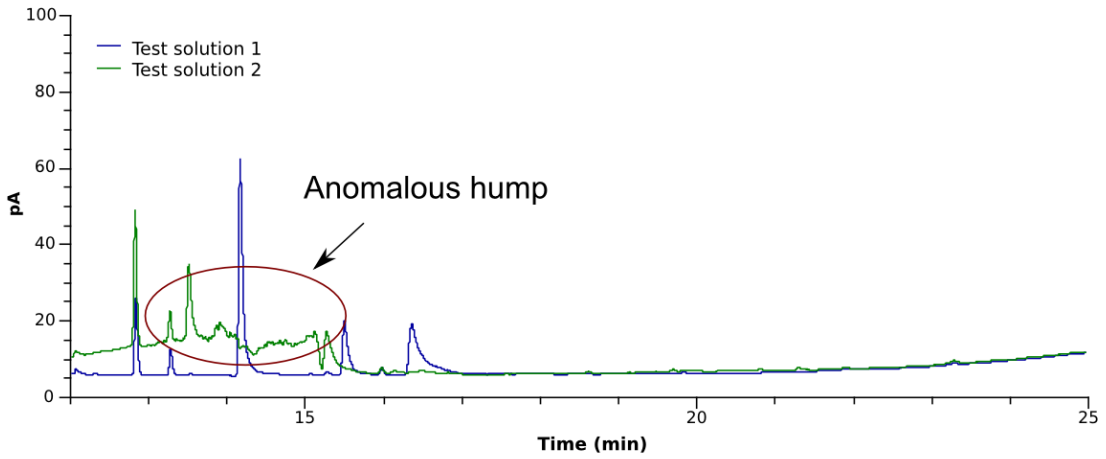


Fig 4.38 GC-FID chromatograms for DMF extracts of aqueous test solutions exhibiting an anomalous hump for the Test solution 2. GC-FID analyses were done within few hours of difference to avoid changes in analysis conditions.

In spite of the fresh sorbents in the SPE column, the GC-FID chromatograms of the calibration curves again exhibited the hump (Fig. 4.39a). Some of the extracts in DMF were analyzed by GC-MS but the nature of the impurity could not be identified as the hump did not appear there (Fig. 4.39b). Based on the presence of the hump in GC-FID, it was assumed to be an organic compound, but as it was not detected by GC-MS, this would suggest that the fragments of a such compound(s) were out of the scan range, i.e. *m/z* lower than 50.

A thorough revision of the analytical steps was made, which included cleaning of glassware and pipeline of SPE apparatus, checking and replacing parts (e.g. installation of a retention gap) of the chromatograph, a refilling of SPE sorbents and preparing fresh aqueous standard solutions and new calibration curves as well.

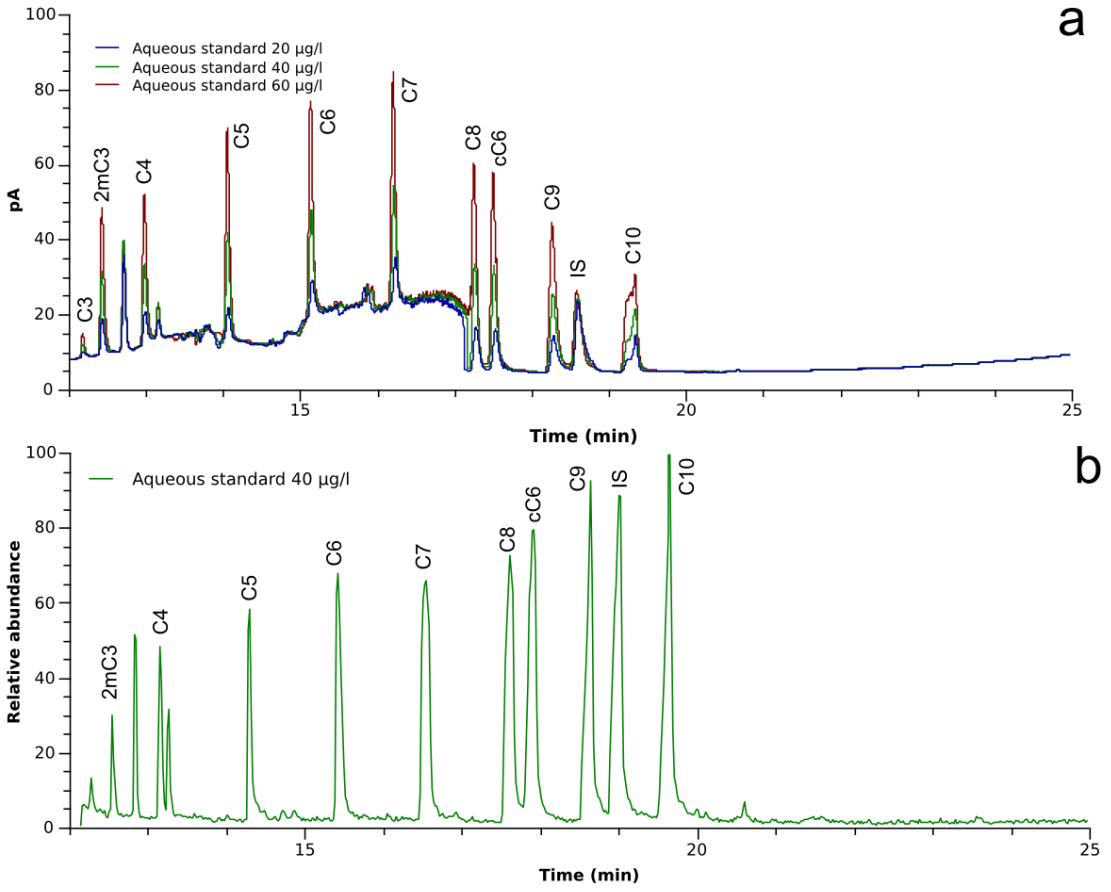


Fig. 4.39 a) GC-FID chromatograms of DMF extracts exhibiting an anomalous hump. b) GC-MS analyses were also made on the same organic extracts but the hump was not obvious in the chromatograms obtained.

It was assumed that the hump is due to a by-product being produced during GC-FID analysis of DMF extracts. The remains of HCl were related to this as DMF may react with HCl to form formic acid and dimethylamine: this reaction can be catalyzed by high temperature. Both the resulting compounds have a molecular weight below 50 g/mol (46 and 45 respectively) preventing their detection in GC-MS.

To remove the excess of HCl, a cleansing step of the SPE column after injection of the sample was tested again. This time, Millipore water and 1M $NaHCO_3$ solution were tested in different volumes of 10, 5, 2, 1 and 0.2 ml. The volume of 200 µl of Millipore water provided good GC-FID chromatograms with no detectable by-products (Fig. 4.40). Thus, the improved SPE procedure now involved: injection of sample, 200-µl rinse with deionized water, drying and elution with acidified DMF (10% formic acid).

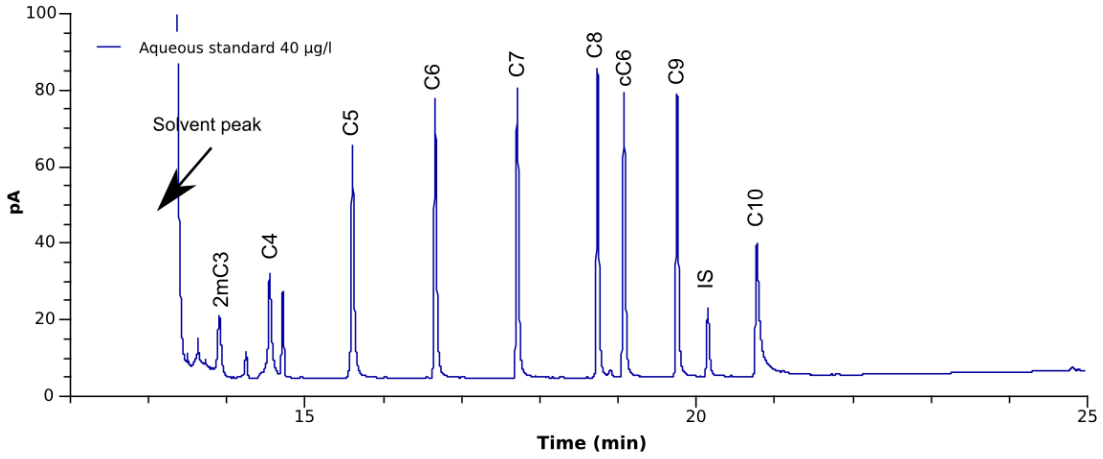


Fig. 4.40 GC-FID chromatogram of SPE extract in acidic DMF which does not show detectable by-products. The improved SPE procedure included a 200-µl rinse with deionized water.

4.3.5 Selection of new internal standard

Due to the different behavior of alcohols and fatty acids in GC and SPE, the internal standard (2-tert-butyl-4-methylphenol) proposed by Jurado-Sánchez et al. (2010) should be substituted by a carboxylic acid. An internal standard must show similar behavior as the target analytes and should neither interfere with the analytes nor be present in natural samples (Cuadros-Rodríguez et al. 2001).

With this in mind a few fatty acids were tested as internal standard (IS), namely 2-ethylbutanoic, cyclopentylpropanoic, cyclohexylpropanoic, 2,4-hexadienoic and 4-ethyloctanoic acid. A DMF standard solution was prepared with these acids and analyzed by GC-FID. As an attempt to assess possible coelution with target analytes, the standard solution also included undecanoic, cyclopentanecarboxylic, cyclohexanecarboxylic and cycloheptanecarboxylic acid.

From Fig. 4.41a the compounds that coelute can be observed partly: heptanoic - cyclopentanecarboxylic acid; octanoic, cyclopentylacetic and cyclohexanecarboxylic acid; and decanoic – cycloheptanecarboxylic acid. Two of the proposed compounds for internal standard exhibited coelution, i.e. 4-ethyloctanoic and cyclopentylpropanoic acid (RT = 18.8 min).

The peak area of 2,4-hexadienoic acid was smaller than the ones for octanoic and nonanoic acid and may indicate partial decomposition. For this acid, thermal decomposition has been reported (Sigma-Aldrich, n.d) above its boiling point (170 °C); during GC-FID analysis the injector temperature rises up to 250 °C. 2-ethylbutanoic acid was also discarded because it eluted close to pentanoic acid, which may interfere with compounds found in extracts of natural samples, as observed when the SPE procedure was tested on a natural sample and the GC-FID chromatogram showed two significant peaks partly coeluting with pentanoic acid (Fig. 4.41b). Cyclohexylpropanoic acid was also excluded as IS because of its high boiling point (275.8 °C), which may risk an effective quantification.

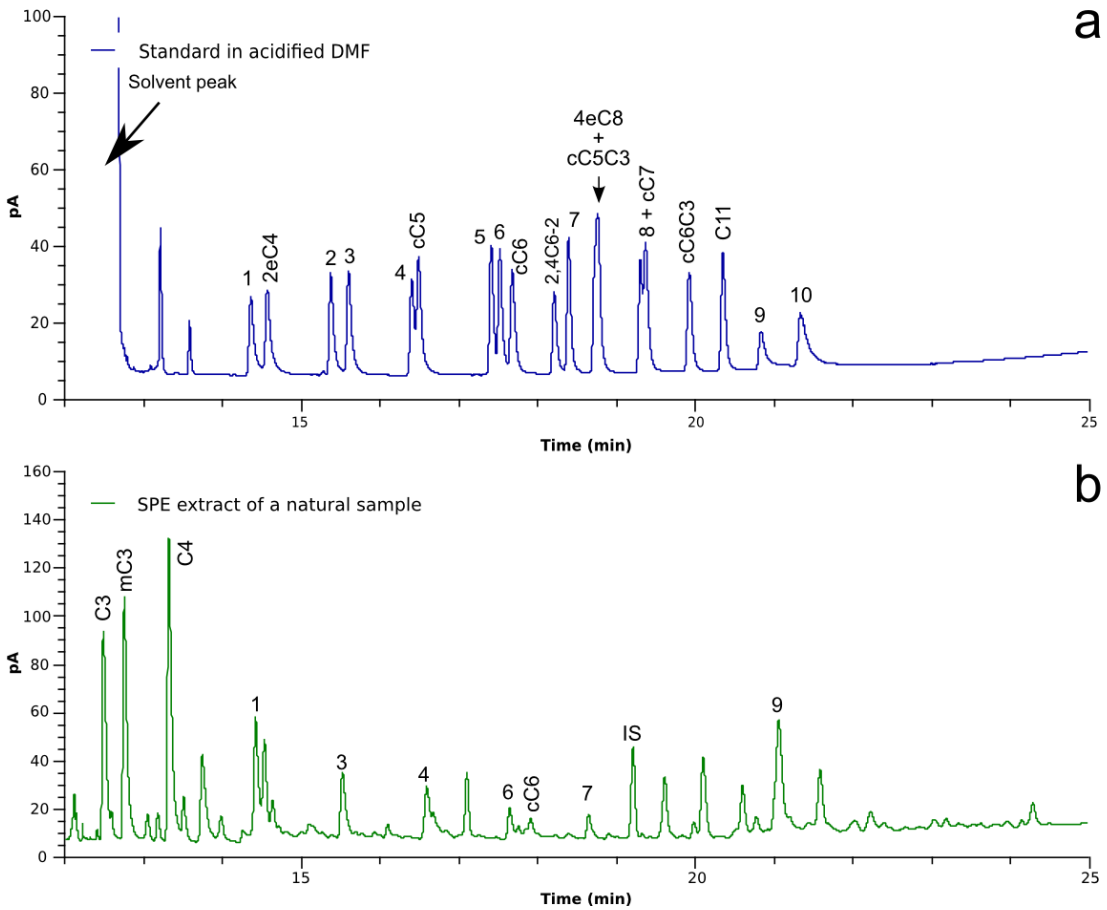


Fig 4.41 Assessment of coelution in the selection of the internal standard. a) GC-FID chromatogram of a standard in DMF containing target fatty acids and the ones proposed as IS, namely 2-ethylbutanoic (2eC4), 2,4-hexadienoic (2,4C6-2), cyclopentylpropanoic (cC5C3), 4-ethyloctanoic, cyclohexylpropanoic acid. b) GC-GFID chromatogram of a SPE extract of a natural sample in DMF. C3, propanoic; mC3, methylpropanoic; C4, butanoic; cC5, cyclopentanecarboxylic; cC6 cyclohexanecarboxylic acid; cC7 cycloheptanecarboxylic acid; C11, undecanoic acid and IS, 2-tert-butyl-4-methylphenol. For nomenclature, see Fig. 4.29.

4-ethyloctanoic has been found in essential oils of costus root, patchouli and olibanum (de Rijke et al. 1978) and its biodegradation is still under investigation (Watson *et al.*, 2002). Wilkes *et al.*, (2003) reported the formation of cyclopentylpropanoic acid in extracts of oil-grown cultures during the degradation of saturated hydrocarbons under anoxic conditions. Thus 4-ethyloctanoic was selected as IS because it can be assumed that its presence is less probable in natural samples.

4.3.6 Optimized SPE procedure

The optimized procedure starts with the acidification of the aqueous extract with 5M HCl to pH below 2 to protonate the target analytes. 50 ml of sample are injected to the SPE column with a flow rate of 2 ml/min followed by 30 ml of air at a flow of 3 ml/min to remove the water excess from the column. Then the direction of the flow is inverted and 200 µl of millipore water are injected to eliminate the HCl remains along with 50 ml of air blown at 3 ml/min to dry the column. The final step is the elution of analytes with 200 µl of acidified DMF at 3 ml/min (Fig. 4.42). Thus, SPE extracts are ready to be analyzed by GC-FID and GC-MS, for quantification and compound identification respectively.

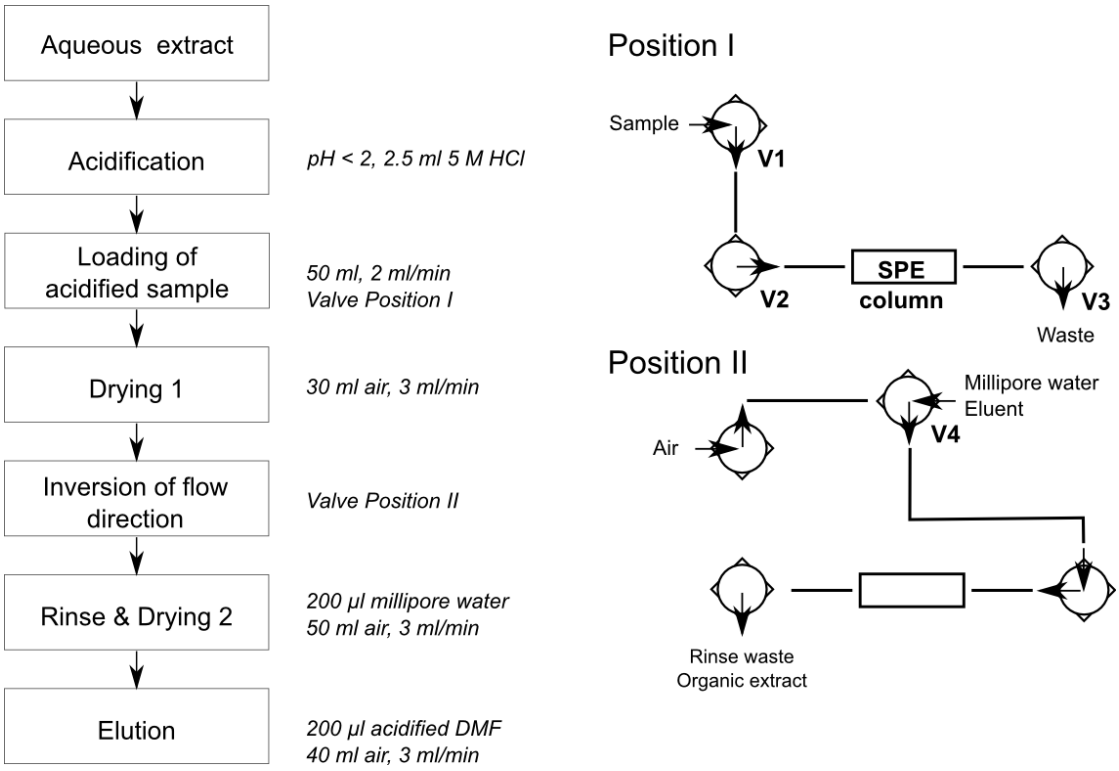


Fig. 4.42 Optimized SPE procedure. For explanation see text.

During the optimization, several kinds of organic acids were assessed. For the final version of the SPE procedure the aqueous standard solutions contained 10 acids: eight

linear (propanoic, butanoic, pentanoic, hexanoic, heptanoic, octanoic, nonanoic and decanoic acid); one branched (2-methylpropanoic acid), and one cyclic (cyclohexanecarboxylic acid), and 4-ethyloctanoic acid was used as IS.

Summary

In the implementation of a method to characterize underivatized LMWOA, acid-base LLE and SPE were tested. The former was discarded due to low recoveries, whereas the latter brought a variety of challenges (e.g. degradation of organic extract) that were successfully overcome. The developed SPE procedure can be applied to determine the amount of saturated LMWOA with a chain length between 5 and 9 carbon atoms contained in aqueous solutions. Thus, this will help to characterize a wider range of water-soluble acids that are released from organic-rich sediments.

5. Characterization of DOC in aqueous extracts from organic-rich sediments

In order to assess the release of organic carbon into the aqueous matrix, size-exclusion chromatography (SEC) was applied to the water extracts. Results will be shown in terms of mg/g-TOC. Subsequently, the correlation between DOC, TOC, and other properties will be discussed.

5.1 DOC and its fractions

DOC in aqueous extracts of coals ranged from 0.62-14.99 mg-C/g-TOC, and the extracts of shales varied between 0.24 up to 10.58 mg-C/g-TOC (Table 5.1). In case of coals, the aqueous extract of Karapiro sample exhibited the highest amount of DOC (14.9 mg-C/g-TOC) followed by the extracts of Gore LM (8.7 mg-C/g-TOC), Lausitz (6.6 mg-C/g-TOC), Waikato (4.2 mg-C/g-TOC), Waikato-1 (1.8 mg-C/TOC), Paparoa (0.7 mg-C/g-TOC) and Cannel (0.62 mg-C/g-TOC).

Table 5.1 DOC in aqueous extracts

Sample	TOC (wt%)	DOC µg-C/g-sed	DOC mg-C/g-TOC
Botneheia	2.17	229.7	10.6
Schöneck	10.1	580.9	5.8
Alum	11.2	27.6	0.25
Posidonia	11.4	432.8	3.8
Irati	25.3	265.9	1.1
Karapiro	45.4	6800.4	14.9
Gore LM	51.6	4503.7	8.7
Waikato-1	60.4	1137.3	1.8
Waikato	64.9	2754.5	4.2
Cannel coal	67.9	424.8	0.62
Lausitz	69.4	4583.7	6.6
Paparoa	74.2	522.5	0.7
OS	12	217.3	1.5
MFT	15.4	0.22	1.4
TS	3.75	3088.4	82.3

DOC in the water extract of Botneheia sample was the highest among shales (10.6 mg-C/g-TOC) followed by those of Schöneck (5.8 mg-C/g-TOC) and Posidonia (3.8 mg-C/g-TOC). The extract of Irati shale extract exhibited a DOC concentration of 1.1 mg-C/g-TOC and the oldest sample, Alum shale (Cambrian-Furongian) showed the lowest concentration (0.25 mg-C/g-TOC). The DOC concentrations in the extracts of

the oil-related samples were 1.5, 1.4 and 82.3 mg-C/g-TOC for the oil sand, MFT and TS respectively. The high value for TS is remarkable, but one must bear in mind all processes (e.g. dewatering and mixing with peat-mineral mix) that this sample has undergone.

Using SEC, DOC is divided into hydrophobic organic carbon (HOC) and hydrophilic organic carbon (CDOC). From Fig. 5.2 it can be seen that water extracts of shales exhibited HOC amount between 25-52 %, and those of coals ranged from 3.5 % to 32 %. The HOC fraction was the highest in the aqueous extract of Botneheia sample (52.4 %) followed by that of Posidonia shale (41.2 %). The extracts of Schöneck and Irati shales showed similar hydrophobic organic matter content, 29.5 and 25.1 % respectively, whereas this fraction was practically absent in the extract of Alum shale, likely due to the recalcitrance of the organic matter in this sample (high maturity), because of which only a low amount of water-soluble organic matter could be extracted.

The HOC content was the lowest in the extract of Waikato-1 coal (3.6 %) and was very similar between the extracts of Karapiro (9 %), Waikato (9 %), and Gore LM (9.7 %) coals; then it increased from 11.3 % in the extract of Lausitz coal up to 32 % in those of Paparoa coal and Cannel Coal. In the oil sand-related samples, the aqueous extract of the TS showed the lowest HOC (2.4 %), with 28.6 % in the MFT, and 87.8 % in the water extract of OS, suggesting the presence of long-chained hydrocarbon residuals in the oil sand.

Parallel to the classification of aquatic hydrophilic organic matter by Huber et al., 2011, the hydrophilic organic carbon in the water extracts could be fractionated according to its chemical functional complexity, molecular weight, and degree of hydrophilicity. In the present work, the heaviest fraction will be labelled as **HMW** (heavier than 20,000 Da), which does not have UV response due as it likely comprises saturated-polymeric compounds; followed by the chemically most complex and recalcitrant fraction named **CoRe** (~1000 Da); the next subdivision is the intermediate fraction (**INT**) the weight of which varies between 300 and 500 Da; finally **LMW-acid** and **LMW-neutral** fractions representing organic matter with MW lower than 350 Da.

The percentage of CoRe fraction ranged from 7.5 up to 33.8 % in the extracts of shales, with those of Botneheia, Schöneck and Posidonia samples exhibiting similar ratios (33.8, 32.4 and 30.5 % respectively), whereas that of Irati shale showed the lowest ratio (7.5 %). The extract of Alum sample showed a CoRe ratio of 16.7 % (Fig. 5.2).

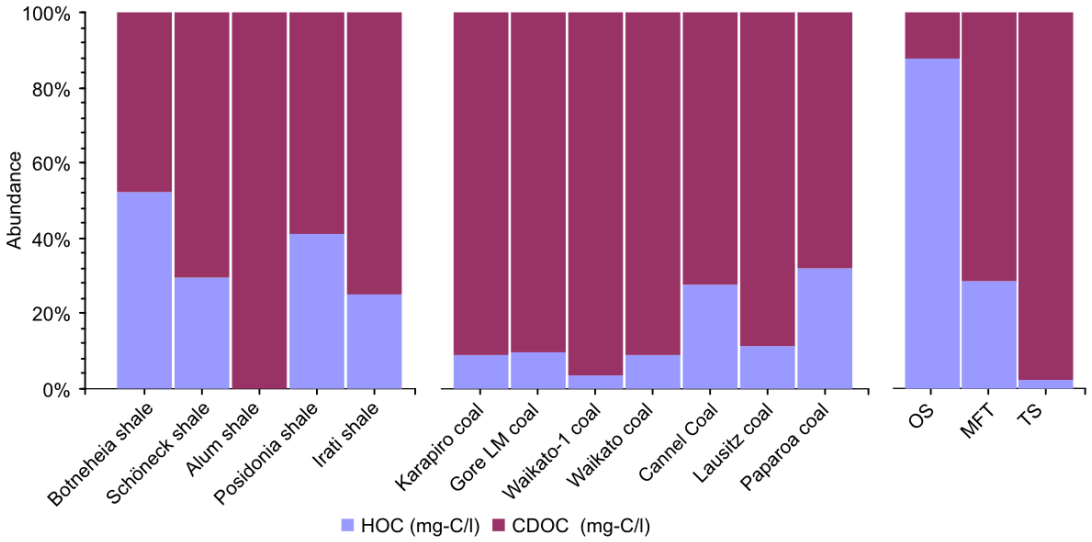


Fig. 5.1 Hydrophobic and hydrophilic dissolved organic carbon in the aqueous extracts of shales, coals and oil sand-related samples.

In the extracts of coals, CoRe fraction reached up to 64 %. The lowest ratio was found in the extract of Cannel Coal (6.7 %), followed by that of Paparoa (11.8 %), Waikato-1 and Waikato (22.1 %), and Lausitz (29.6 %) extracts. The extracts of Karapiro and Gore LM samples presented the highest percentage of CoRe ratios, 64 and 51.2 % respectively. The extracts of OS and MFT exhibited 20.5 and 26.7 % CoRe fraction, while TS extract having the highest value at 75 %.

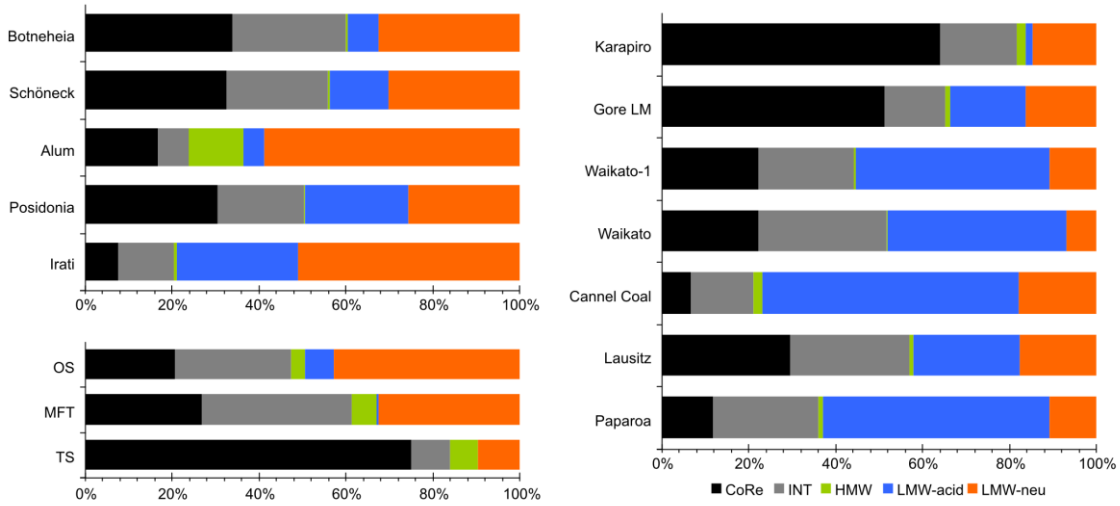


Fig 5.2 Distribution of fractions of organic matter regarding molecular weight, chemical complexity and hydrophilicity by LC-OCD in the water extracts of shales, coals and oil sand-related samples.

INT fraction amounted to 26.1 and 23.5 % in the extracts of Botneheia and Schöneck shales, respectively; those of Posidonia shale and Irati Fm showed 19.9 and 12.9 % INT fraction, while in the extract of Alum this fraction represented just 7%. Among coals, the extract of Waikato exhibited an INT fraction of 29.6 %, followed by the extracts of Lausitz (27.3 %), Paparoa (24.3 %) and Waikato-1 (22.1 %). This fraction was 17.7, 14.5 and 13.9 % in the extracts of Karapiro, Cannel Coal and Gore LM respectively. In the extract of MFT this fraction reached 34.5 % and 26.8 % in that of OS, and it was just 8.9% in the extract of TS.

The HMW fraction was low in the extracts of most coals and shales, ranging from 0.24 up to 2.2 %, but it varied between 3.3 up to 6.4 % among those of oil sand-related samples. An exception was the extract of Alum shale with 12.4 % of HMW.

In shales the LMW-acid fraction increased as follows: 4.6 % (Alum extract); 7% (Botneheia extract), 13.6 % (Schöneck extract), 23.7 % (Posidonia extract) and 27.8 % (Irati Fm extract). The extract of Karapiro sample showed a remarkably low LMW-acid fraction (1.4%), whereas of Gore LM reached 17.4 %. The Lausitz extract was 24.5 % followed by that of Waikato and Waikato-1 with a LMW-acid fraction of 44.6 and 41 % respectively. The extracts of Paparoa and Cannel coal exhibited the highest ratios, 52.2 and 59.1 % respectively. In the oil sand-related samples, this fraction reached 6.7 % in the in the extract of the OS and 0.5 % in the MFT extract.

Regarding the LMW-neutral fraction, in shales the highest percentage was found in the extract of Alum sample (58.9 %) followed by that of Irati Fm (51.1 %). The extracts of Botneheia and Schöneck samples showed 32.5 % and 30.2 %, respectively; this fraction in the extract of Posidonia shale was 25.6 %. In the extracts of coals, the LMW-neutral increased as follows: 7 % (Waikato), 10.8 % (Waikato-1 and Paparoa), 14.7 % (Karapiro), 16.2 % (Gore LM), 17.6 % (Lausitz) and 17.9 % (Cannel coal). In the extracts of oil-related samples, the extract of TS exhibited a neutrals fraction of 9.7 %, in comparison to that of OS which reached a high of 42.7 %. In the extract of MFT this fraction accounted for 32.4 %.

The DOC fractions data show that the HMW fraction was low in the extracts of most coals and shales, but was higher among those of oil sand-related samples. An exception was the extract of Alum shale, but considering that the HMW fraction is related to high molecular weight organic matter, the maturity of this sample (the highest amongst all) may be reflected on the higher proportion of these compounds. However, biological activity may also contribute to an increase of these compounds as exhibited by OS likely as microbial matter, and the tailing sand which has already undergone plant growth implying an active rhizosphere.

The terrestrial origin of coals may be responsible for higher CoRe fraction in comparison to the marine organic matter source of shales. This fact is probably supported by the amount of this fraction in the tailings sand, which exhibited the highest proportion of CoRe, TS has been enriched with peat and has also experienced plant growth. LMW-neutrals basically are alcohols, aldehydes and ketones. Amino acids can also elute in this fraction due to their low net charge. The predominance of LMW-neutrals in shales could be due to the type of organic matter related to a more reactive and protein-rich aquatic environment than the terrestrial one. It may also result from microbiological activity as suggested by oil-related samples, in which LMW-neutrals was also predominant over the LMW-acids.

As it will be shown later (See Chapter 6), generally the proportion of LMW-acids agreed with the amount of carboxylic acids quantified by IC, namely carboxylic content of coals

was much higher than that of shales. However, the cases of the extracts of TS in which LMW-acids were not detected, and Waikato-1 in which this fraction was very small, need to be clarified.

5.2 Relationships between DOC, TOC, maturity and kerogen type

In order to evaluate the relationships between DOC and some properties of the samples (e.g. TOC), Pearson correlation coefficient was calculated in terms of DOC as mg-C/g-TOC. Because every oil-sand related sample represents a particular environment, they were not included in this calculation. From Table 5.1 it can be observed that DOC, quantified as mg-C/g-sed, was higher in the extracts of coals than those of shales and oil-sand related samples, which may be related to their higher TOC content. The Student's test was used to prove the null hypothesis - the means of DOC (mg-C/g-sed) are equal in extracts of shales and coals. For a two-tailed test, with a P value of 0.04 (degrees of freedom = 10, t-value = 2.4) the null hypothesis is rejected. This suggests that either TOC or organic matter type influences the DOC in the extracts. However, when Pearson's coefficient was calculated, DOC (mg-C/g-sed) did not correlate with TOC (r = 0.42, R^2 = 0.18) in the extracts of shales and coals.

Table 5.2 Pearson correlation coefficient of DOC (mg-C/g-sed) with some sample properties

	Shales		Coals	
Sample	**r**	**R^2**	**r**	**R^2**
TOC (wt%)	-0.02	0.00	-0.73	0.53
HI	0.72	0.52	-0.37	0.14
OI	0.55	0.31	0.82	0.67
Tmax	-0.75	0.56	-0.71	0.51
Ro	-0.75	0.56		
Sr			**-0.87**	**0.75**

New correlation Pearson's coefficients were then calculated based on the sample type (Table 5.2). In case of the aqueous extracts of shales, the absence of correlation between DOC and TOC (r = -0.02, R^2 = 0.00) remained, whereas DOC in the coal water extracts was found to be weakly negatively correlated with TOC (r = -0.73, R^2 = 0.53). The absent of correlation of DOC with TOC may be related to the less complex sediment structure and marine origin of the organic matter of the shales. For instance,

kerogen type I contains more long-chain aliphatics in comparison to kerogen type III, whereas the latter more aromatic moieties. This can be interpreted as the structure of organic matter in coals is more intricate and tightly bond than in the shales; which hinder the hydrolysis of more water-soluble molecules. This is supported by a Student's test applied to the DOC in terms of mg-C/g-TOC. A P-value equal to 0.72 (t-value = 0.37) shows that the means of DOC in the extracts of shales and coals are not significantly different.

Nevertheless DOC in the extracts of shales exhibited a weak positive correlation with HI ($r = 0.72$, $R^2 = 0.52$), this parameter did not correlate in case of coal extracts ($r = -0.37$, $R^2 = 0.14$). An opposite behavior was observed regarding OI, Pearson coefficient showed a weak positive correlation between DOC and OI ($r = 0.82$, $R^2 = 0.67$, whereas no correlation was found in case of extracts of shales ($r = 0.55$, $R^2 = 0.31$).

Different maturity parameters were used - Tmax as indicator of thermal maturity for both shale and coal extracts; vitrinite reflectance in case of shales, and Suggate rank for the coal samples. In the extracts of shales and coals, weak negative correlations were found between DOC and Tmax, $r = -0.75$, $R^2 = 0.56$ and $r = -0.71$, $R^2 = 0.51$ respectively. In case of shales, Pearson's r showed a weak negative correlation between DOC and vitrinite reflectance, $r = -0.75$, $R^2 = 0.56$; whereas DOC correlated with Suggate rank strongly and negatively in coal samples (r = -0-87, $R^2 = 0.75$). These result accord with the absent of correlation between DOC and TOC, as more mature samples represent more condensed structures, then there are few reactive sites for hydrolysis.

In addition, the absence or weakness of correlation may be interpreted as no clear predominance of any of these parameters over the release of the fraction that is dissolved into the aqueous matrix and quantified as DOC. An exception being made, in case of the coals extract in which maturity (Suggate rank) showed a significant correlation with DOC.

From Fig 5.3 (left), it can be seen that the extracts of Paleogene sediments exhibit DOC concentrations similar to those of older samples, e.g. Botneheia (10.6 mg-C/g-

TOC) and Gore LM (8.7 mg-C/g-TOC). Further, differences between recent sediments were observed, for instance, the DOC in the extract of Waikato-1 was much smaller than that in the extract of Gore LM. The DOC in the extract of Irati shale, from Permian, was slightly higher that than in the extract of Paparoa from Cretaceous. Thus, within the analyzed samples no evident relationship between DOC in extracts and geological age was observed. In a HI - Tmax diagram (Fig. 5.3), it can be seen that extracts of samples identified as kerogen type I and II exhibited low and medium amounts of DOC expressed as mg-C/g-TOC, whereas in aqueous extracts of samples of kerogen type III, DOC varied from low to high values.

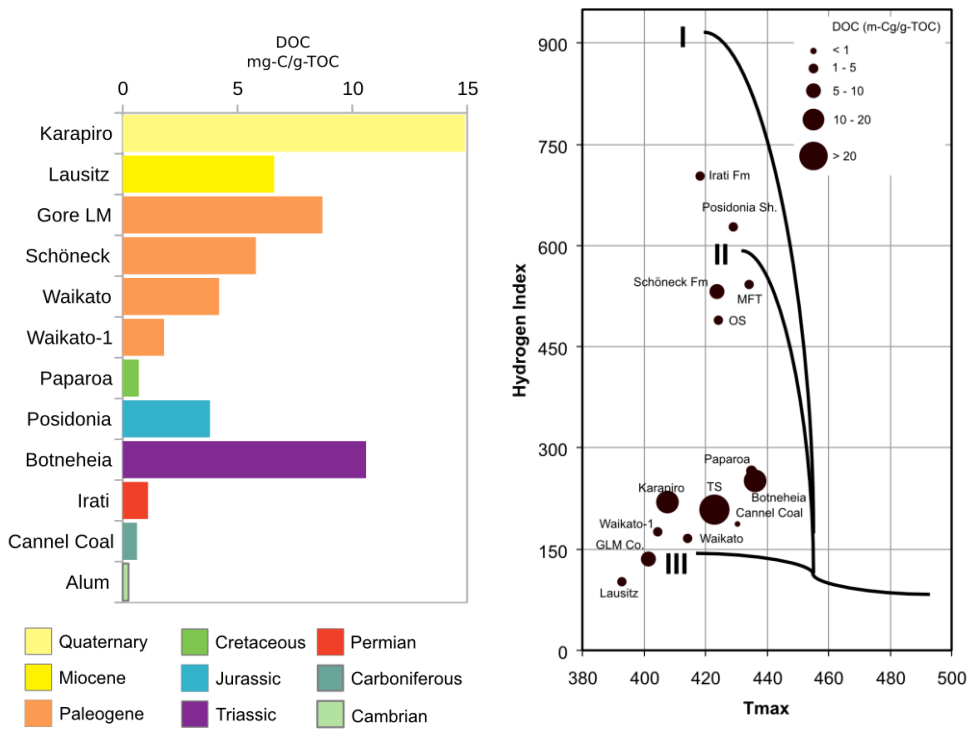


Fig 5.3 Dissolved organic carbon (mg-C/g-TOC) in water extracts of natural samples in a plot bar with geological age (left); and, a HI (mg-HC/g-TOC) vs Tmax (°C) plot (right).

Summary

Characterization of the water extracts of shales and coals was performed by analysis of DOC and its fractions and identification water soluble polar compounds. DOC (mg-C/g-sed) in the extracts of coals was larger than that in the shale extracts. However, when DOC was normalized to TOC and used for calculations, a Student's test showed that the means of DOC in the extracts of shales and coals were not significantly different, suggesting that other properties play a role as well. Pearson's correlation coefficient showed that DOC (expressed as mg/g-sed) weakly negatively correlated with TOC, and a weak positive correlation between DOC and OI was observed. In the extracts of shales and coals, weak negative correlations were found between DOC and Tmax. In the extracts of shales, Pearson's r showed a weak negative correlation between DOC and vitrinite reflectance; whereas, DOC correlated with Suggate rank strongly and negatively in the extracts of coals.

6. Characterization of water-soluble LMWOA that are released from organic-rich sediments

With a strategy based on IC and SPE-GC-FID/GC-MS, this study provides a quantitative characterization of water-soluble LMWOA with chain length between 1 and 9 carbon atoms, released from organic-rich sediments, that can serve as feedstock for the deep biosphere. For understanding the factors that are linked to the release of the acids, the relationships between acid content, kerogen type, geological age, TOC, maturity and inorganic ions were examined. Further, to know to what extent the yields of water-soluble LMWOA may be affected, Waikato-1 coal and Posidonia shale were tested under two extraction conditions: (i) three different water/sediment ratios, by varying the amount of sediment, specifically 500:1, 100:1 and 50:1 ratios; and, (ii) pH and phosphate content, the extractions were done with 5 g of sediments in aqueous buffer solutions of pH 2, pH 7, and 12.5.

6.1 Polar compounds dissolved in water extracts

The techniques employed (i.e. IC and SPE) were focused on acidic and moderately acidic compounds that allowed the identification of around 36 different substances (Table 6.1) as phenols, phenolic aldehydes and aliphatic, branched, aromatic, dicarboxylic, hydroxy and keto acids from one carbon chain length up to ten in the water extracts. Fig. 6.1 presents GC-MS chromatograms of two samples, Schöneck shale and Gore LM coal, showing the diversity of such compounds in the samples.

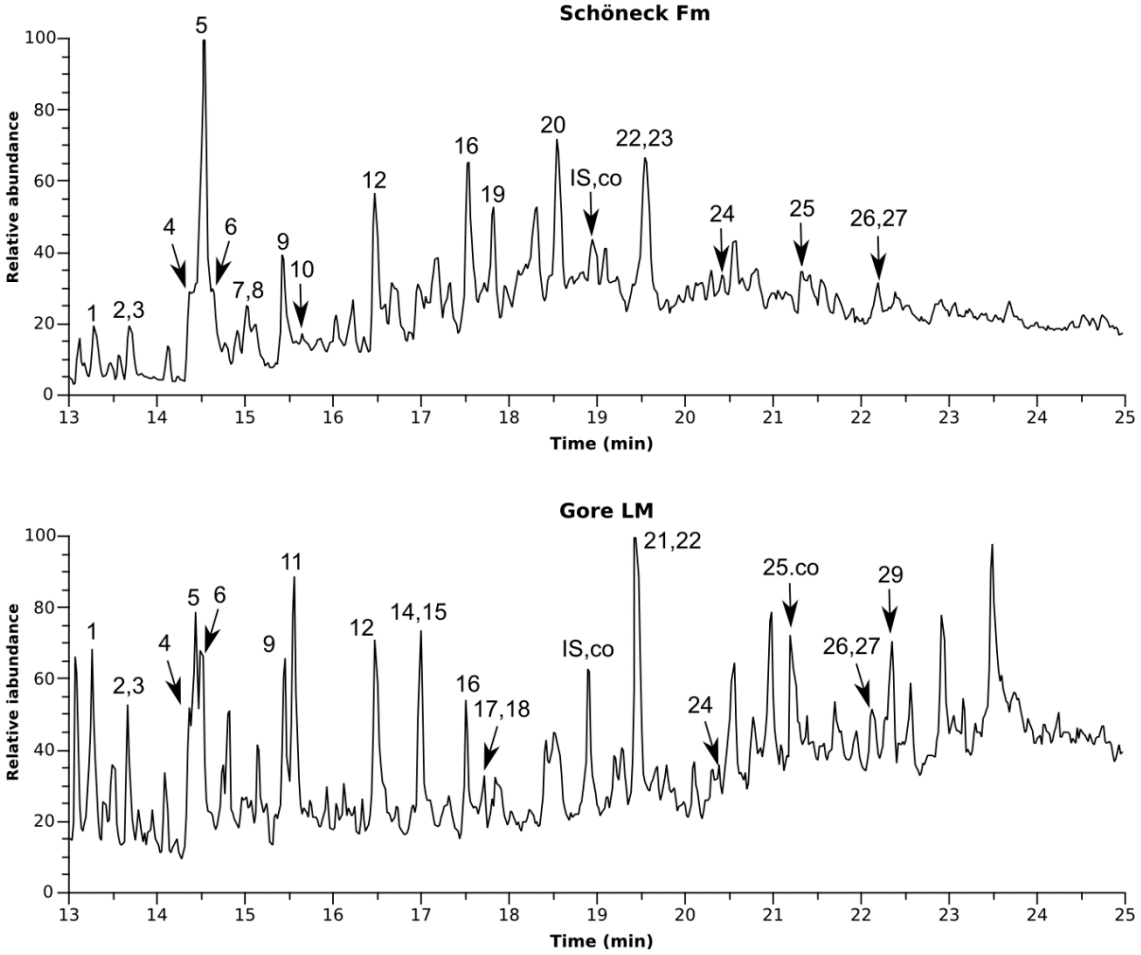


Fig. 6.1 GC-MS chromatograms of SPE extracts of Schöneck shale and Gore LM coal. For nomenclature see Table 6.1

Table 6.1 Compounds identified in the aqueous extracts by IC and GC-MS

		Shales					Coals							Oil sand related samples		
Compound	Code	Botneheia	Schöneck	Alum	Posidonia	Irati Fm	Karapiro	Gore LM	Waikato-1	Waikato	Cannel Co	Lausitz	Paparoa	OS	MFT	TS
TOC (% wt)		2.17	10.1	11.2	11.4	25.3	45.38	51.61	60.41	64.92	67.9	69.4	74.18	12	15.4	3.75
Formic acid[1]																
Acetic acid[1]																
Propanoic acid[1]																
Butanoic acid[1]																
Present																
Coelution		Co														

1, these compounds were identified/quantified by IC. 2, butanoic acid was identified by IC and SPE.

Table 6.1 Compounds identified in aqueous extracts by IC and GC-MS. Continuation

		Shales					Coals							Oil sand related samples		
Compound	**Code**	**Botneheia**	**Schöneck**	**Alum**	**Posidonia**	**Irati Fm**	**Karapiro**	**Gore LM**	**Waikato-1**	**Waikato**	**Cannel Co**	**Lausitz**	**Paparoa**	**OS**	**MFT**	**TS**
Pyruvic acid[1]																
Methylsuccinic acid[1]																
Malic acid[1]																
Oxalic acid[1]																
Butanoic acid[2]	1															
2-Methylbutanoic acid	2		Co		Co	Co	Co	Co	Co	Co	Co	Co	Co			
3-Methylbutanoic acid	3		Co		Co	Co	Co	Co	Co	Co	Co	Co	Co			
Pentanoic acid	4															
2,3-dimethylmaleic anhydride (DMMA)	5	Co	Co		Co							Co				
Unknown	6		Co													
2-methylpentanoic acid	7				Co	Co				Co			Co			
4-methylpentanoic acid	8	Co	Co		Co	Co				Co			Co			
Hexanoic acid	9															
2-methylhexanoic acid	10															
2-methoxyphenol (Guaiacol)	11															
Heptanoic acid	12															
Thymoquinone (tent.)	13				Co											
o-cresol	14					Co	Co	Co		Co	Co	Co	Co	Co	Co	
Phenol	15			Co		Co	Co	Co	Co	Co	Co	Co	Co	Co	Co	
Octanoic acid	16															
p-cresol	17					Co	Co	Co	Co	Co	Co	Co	Co			
m-cresol	18					Co	Co	Co	Co	Co						
Cyclohexanecarboxylic acid	19															
Nonanoic acid	20															
4-ethyloctanoic acid (IS)	*IS*	*Co*	*Co*		*Co*	*Co*	*Co*	*Co*	*Co*							*Co*
2,6-dimethoxyphenol (Syringol)	21						Co	Co			Co					
2-ethyl-3-methylmaleimide (EMMI)	22		Co				Co	Co			Co					
Decanoic acid	23	Co	Co													
2-hydroxymethylbenzoic acid	24															
Benzoic acid	25					Co										
3-methylbenzoic acid	26	Co	Co	Co	Co	Co	Co	Co		Co	Co		Co			Co
4-methylbenzoic acid	27	Co	Co	Co	Co	Co	Co	Co		Co	Co		Co			Co
Vanillin	28															
Phenylacetic acid	29															
F-																
Cl-																
NO_3^-																
SO_4^{2-}																
PO_4^{3-}																
Present																
Coelution		Co														

Amongst shales, the aqueous and SPE extracts of Alum shale exhibited a diversity of 15 compounds, followed by those of Svalbard shale (24 acids). In the former only some monocarboxylic and aromatic acids were identified, while in the extracts of the latter oxalic acid and some methyl-branched acids were also found. The composition of extracts from Schöneck (29 acids), Posidonia (27 acids) and Irati (31 acids) shales was similar but with some subtle differences. Oxalic acid was present in the organic extracts of these three samples, pyruvic acid was detected only in Irati extract, and methylsuccinic acid was only detected in Schöneck extract. Thymoquinone, an aglycone[1] found in some trees and plants, was tentatively identified in the SPE extract of Posidonia shale[2]. Thymoquinone was not the only aglycone found; in organic extracts of coals guaiacol was also present.

Aliphatic monocarboxylic acids from formic up to octanoic acid were identified in the extracts of most coals, with some exceptions. Heptanoic and octanoic acid were not detected in the extract of Cannel coal. Nonanoic acid was only found in the SPE extract of Waikato-1 coal. Pyruvic, oxalic and methylsuccinic acid were present in the the extracts of coals, except in that of Paparoa coal where the latter was no identified. Unlike those of shales, all extracts of coals exhibited malic acid. The extract of Paparoa, the most mature coal with the highest TOC, exhibited the lowest diversity of compounds. With respect to branched acids, 2 and/or 3-methylbutanoic acid were also found in all coals; however methylpentanoic acids were present just in the extracts of Waikato and Paparoa coals. Unlike the extracts of coals, those of the shales exhibited the presence of decanoic acid, but its relative abundance was high in the extract of Posidonia shale. As will be shown later, this sample was prone to release even-chained carboxylic acids. Apart from the dicarboxylic acids quantified by IC, no other short-chained dicarboxylic acids were detected in the SPE extracts. It is not clear whether they were present or absent in the SPE extracts as, many of these acids decompose at high temperatures, similar to those required for GC analysis.

[1] When a hydrogen atom replaces the glycosyl group from a glycoside, the product is called aglycone as opposed to glycone which correspond to the sugar group. Glycosides are a class of molecule in which, a sugar molecule (gylcosyl) is bonded to a “non-sugar” molecule by a glycosidic bond.

[2] Identification was done by comparison with the mass spectra library. Verification with a standard in DMF failed because thymoquinone underwent a reaction during the standard preparation or GC-MS analysis, as a result the obtained mass spectrum was different from the one in the MS library.

Phenol was more frequently observed in coals. *o*-cresol was found in nearly all samples, but it was coeluting with phenol in all coal extracts and that of Irati shale. The other isomers, *m*- and *p*-cresol were also mainly present in coal extracts eluting after octanoic acid. The origin of these compounds is still unclear. While Bracewell *et al.* (1980) reported the generation of phenol and cresols as products of the pyrolysis of polymaleic acid (considered as a model compound representing fulvic acids in soils), Saiz-Jimenez and De Leeuw (1984) did not find these aromatic compounds to be representative in their pyrogram of polymaleic acid.

Bracewell *et al.* (1980) identified dimethylmaleic anhydride (DMMA), detected in the extracts of most samples (Table 6.1 Fig. 6.2), as was one of the two characteristic products of the pyrolysis of polycarboxylic acids. These authors reported that DMMA could have been formed by fragmentation, cyclization, elimination and decarboxylation reactions that polycarboxylic acids undergo during pyrolysis. Additionally, Ikan *et al.* (1986) pointed out that heterocyclic-hydroxy-furanone, pyrrole, furan and hydroxy-cyclopentenone derivatives, such as DMMA, are the main building blocks in the polymeric core structure of humic substances. Larsen *et al.* (2005) reported the formation of carboxylic anhydrides and water from carboxylic acids in isolated kerogens and whole Bakken shales at low temperatures (40 - 180 °C). These authors point out that the observed fast reaction indicates that the carboxylic acids must be adjacent to each other in the sediment.

Lignin-derived components such as 2-methoxyphenol (guaiacol) and 2,6-dimethoxyphenol (syringol) were present in most coal extracts, but it was absent in the extracts of shales. In extracts of Karapiro) and Gore LM coals vanillin, an indicator of terrestrial organic matter was also present.

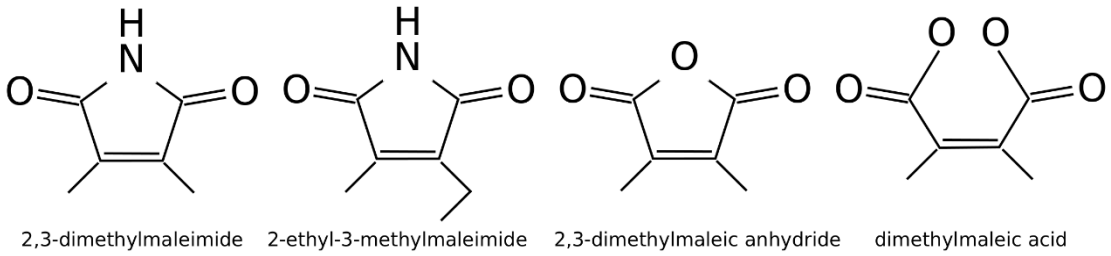


Fig. 6.2 2D structures of some maleimides, anhydride and related acid.

2-ethyl-3-methylmaleimide (EMMI) was detected in extracts of all coals as well as in Schöneck and Irati shale extracts. This finding is interesting, especially as EEMI serves as biomarker for chlorophylls (Grice *et al.*, 1996), which may form under photic and/or oxygen rich sedimentary conditions (Kozono *et al.*, 2002). It was detected in the extract of Schöneck shale which is thought have been deposited in a basin with an oxygen-depleted bottom water. In the extracts of Karapiro, Gore LM and Cannel Coal, this compound was present coeluting with syringol. Recently, Naeher *et al.* (2013) studied naturally occurring maleimides (labeled as free) and those released after chromic acid oxidation of sediment extracts (bound maleimides) in recent sediments from the Swiss lake Rotsee and the Romanian Black Sea Shelf. They identified 6 classes of maleimides, amongst them 2,3-dimethylmaleimide (DMMI) and EMMI (Fig. 6.2); the latter was identified as the predominant homologue in both the forms, free and bound maleimides. These authors also observed that in the Rotsee surface sediment, free maleimides were almost absent and their concentrations increased with depth, which was interpreted as maleimides primarily formed within the sediment, nevertheless the process should still be elucidated.

From a matter-structural point of view, it is worth commenting that an unidentified substance was frequently found in most water extracts, notably in the chromatograms of the extracts of Schöneck and Posidonia shales. In GC chromatograms, this peak was close to those of pentanoic acid and dimethylmaleic anhydride peaks (coeluting sometimes). Based on its mass spectral fragmentation pattern (*m/z* 53, 67, 112, 68, 140), it does not seem to be a maleimide. According to Kozono et al. (2001) the major m/z values of 2-methylmaleimide and 2,3-dimethylmaleimide are 111 and 125 respectively, additionally to 2-ethyl-3-methylmaleimide, with a m/z value of 139. Further studies are needed to identify this compound and to elucidate its significance.

Other aromatic compounds found in most samples were 2-hydroxymethylbenzoic, benzoic, 3-methylbenzoic (*m*-toluic acid) and/or 4-methylbenzoic acid (*p*-toluic acid). The differentiation between isomers of toluic acid was not possible due to coelution. Machihara and Ishiwatari (1983) investigated the stability of compounds that have been detected as degradation products of kerogen by alkaline permanganate (KMnO4) oxidation method at 60 °C. These authors reported that benzoic acid can be produced by oxidation of amino acids (phenylalanine) and proteins.

In oil sand-related samples, formic acid was detected in all three extracts, but acetic and oxalic acids were only present in two of them (OS and TS). Most of the saturated carboxylic acids from 5 up to 10 carbon chain length were also detected in the three aqueous extracts. The extract of MFT also exhibited some methyl branched acids and guaiacol; this may be related to contamination during storage in the tailing pond and later processing (Noah *et al.*, 2014). In the extract of TS, guaiacol, syringol, vanillin and some benzoic acids were also detected, which could be due to the fact that this sample has been mixed with peat and undergone plant growth.

6.2 LMWOA and inorganic ions in the extracts of coals and shales

This section describes concentrations of LMWOA with chain length between 1-9 carbon atoms (Table 6.2 and Fig. 6.3-5), and some inorganic anions in the extracts of organic-rich sediments. The concentrations were normalized to TOC content in order to compare the amounts of released acids. The concentration of individual lineal acids varied amongst samples, and the implications of this variability will be discussed in following sections.

Saturated carboxylic acids

In the shale extracts, formic acid content ranged from 35.2 (Alum) up to 1318 µg/g TOC (Botneheia), whereas acetic acid exhibited concentrations varying between 182.4 in the extract of Alum shale and 1993.3 µg/g TOC in that of Posidonia shale. In case

of coals, the concentration of formic acid in the extract of Cannel Coal exhibited the lowest value (72.1 µg/g TOC) and the Gore LM extract the highest (5470.2 µg/g TOC); while the acetic acid content was between 1163.6 (Paparoa) up to 4472.1 µg/g TOC (Waikato). The content of propanoic acid varied from undetectable up to 270.3 µg/g TOC in the aqueous extracts of shales, and an even higher value of 348 µg/g-TOC in coals. In the extracts from Schöneck, Irati and Posidonia samples, butanoic acid was detected ranging from 11.7 up to 106 µg/g TOC. This compound was also found in extracts of two coals (Gore LM and Waikato), where the concentration varied between 79.4 and 122.5 µg/g TOC. Amongst the oil sand-related samples, only acetic acid was present in concentrations above the detection limit (159.9 µg/g TOC) in the extract of TS.

In the extracts of shales, the concentration of pentanoic acid (*n*C5) varied from 7.3 (Irati) up to 48.2 µg/g-TOC (Schöneck). An exception is the Alum shale extract, in where the concentration of this acid was below the detection limit. In the extracts of coal, that of Cannel Coal exhibited the lowest concentration, 0.9 µg/g-TOC, while the extract of Gore LM showed the highest value (9.4 µg/g-TOC). The concentrations of hexanoic acid (*n*C6) ranged from 0.4 up to 76.8 µg/g-TOC in the extracts of shales, and in those of coals the content of this acid varied between 0.5 and 8 µg/g-TOC. The extracts of shales exhibited heptanoic acid (*n*C7) content from below the detection limit up to 67 µg/g-TOC; while in the coal extracts the concentration of this acid ranged from below the detection limit up to 3.4 µg/g-TOC. Octanoic acid (*n*C8) was not detected in the extracts of Alum Shale, Cannel Coal and Paparoa. However, in the remaining samples, the range of concentration of this acid varied from 1.6 up (Irati) to 66.7 µg/g-TOC in extracts of shales; while the extracts of coals exhibited concentrations between 0.2-1.1 µg/g-TOC. Concentrations of nonanoic acid (*n*C9) ranged from 19.4 up to 63.9 µg/g-TOC in extracts of shales; however, this acid was not detected in the extracts of coals, with exception of the Waikato-1 extract (0.6 µg/g-TOC).

Table 6.2 Concentration of carboxylic acids and some inorganic anions in the water extracts.

Concentration (µg/g-TOC)	Botneheia (1S)	Schöneck (2S)	(3S)Alum	Posidonia (4S)	(5S)Irati	Karapiro (1C)	Gore LM (2C)	Waikato-1 (3C)	(4C)Waikato	(5C)Cannel Coal	(6C)Lausitz	(7C)Paparoa	OS	MFT	TS
TOC	2.17	10.1	11.2	11.4	25.3	45.38	51.61	60.41	64.92	68.9	69.4	74.12	12	15.4	3.75
Formic acid	1318.3	< LOD	35.2	730.6	159.7	2716.6	5470.2	2991.5	2357.3	72.1	3702.5	333.8	< LOD	< LOD	< LOD
Acetic acid	225	1866.2	182.4	1993.3	735.1	1218.8	4296.4	2364.4	4472.1	975.1	1023.1	1163.6	< LOD	n.d.	159.9
Propionic acid	n.d.	270.3	n.d.	129.7	110.3	9.6	184.5	273	348	67.6	60.9	< LOD	n.d.	n.d.	n.d.
Butanoic acid	n.d.	11.7	n.d.	106	25.5	n.d.	122.5	n.d.	79.4	n.d.	n.d.	< LOD	n.d.	n.d.	n.d.
Pentanoic acid	29.5	48.2	< LOD	20.9	7.3	2.9	9.4	4.3	9	0.9	6.1	2.1	3.2	7.8	18.1
Hexanoic acid	25.8	76.8	0.4	64.5	4.5	3	8	1.7	3.7	0.5	4.3	0.7	2.7	6.7	21.5
Heptanoic acid	21.3	67	< LOD	17.6	2.2	1.1	3.4	0.9	1.3	n.d.	1.5	< LOD	2.2	4.6	8.4
Octanoic acid	23	66.7	< LOD	57.1	1.6	0.9	0.5	1.1	0.2	n.d.	1	n.d.	2.7	4.6	19.7
Nonanoic acid	31.8	63.9	< LOD	19.4	n.d.	n.d.	n.d.	0.6	n.d.	n.d.	n.d.	n.d.	n.d.	3.9	n.d.
Oxalic acid	1494.7	752.4	n.d.	582.2	107.8	3539.7	1158.3	1199.9	2736.3	135.7	1710.7	338.9	n.d.	377.8	3811.6
Pyruvic acid	n.d.	n.d.	n.d.	n.d.	95.8	1307.1	884.9	283.1	375.8	27.1	378.2	236.2	n.d.	n.d.	n.d.
Malic acid	n.d	n.d.	n.d.	n.d.	n.d.	322.8	813.6	338.9	377.8	15.7	189.2	< LOD	n.d.	n.d.	n.d.
Methylsuccinic acid	n.d	2255.3	n.d.	n.d.	n.d.	74.6	280.5	136.4	181.6	n.d.	63.4	<LOD	n.d.	n.d.	n.d.
F^- *	10.2	15	36.4	11.2	21.4	21.4	13.7	7.5	10.5	10.7	8.3	11	1	13	1.8
Cl^- *	13.7	22.2	26.1	26.9	9.7	48.1	23.5	8.6	4.4	70.1	14.1	563.4	20	1088.5	8.3
$NO3^-$ *	5.8	9.4	n.d.	9	5	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
$SO4^{2-}$ *	4214.1	7932.8	17357.9	18843.1	9075.8	801.3	1227.2	357.9	68	11256.1	558.6	52	116.6	n.d.	215.4
$PO4^{3-}$ *	n.d.	12.5	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	26	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.

< LOD below limit of detection; n.d. no detected; *concentration as µg/g sed.

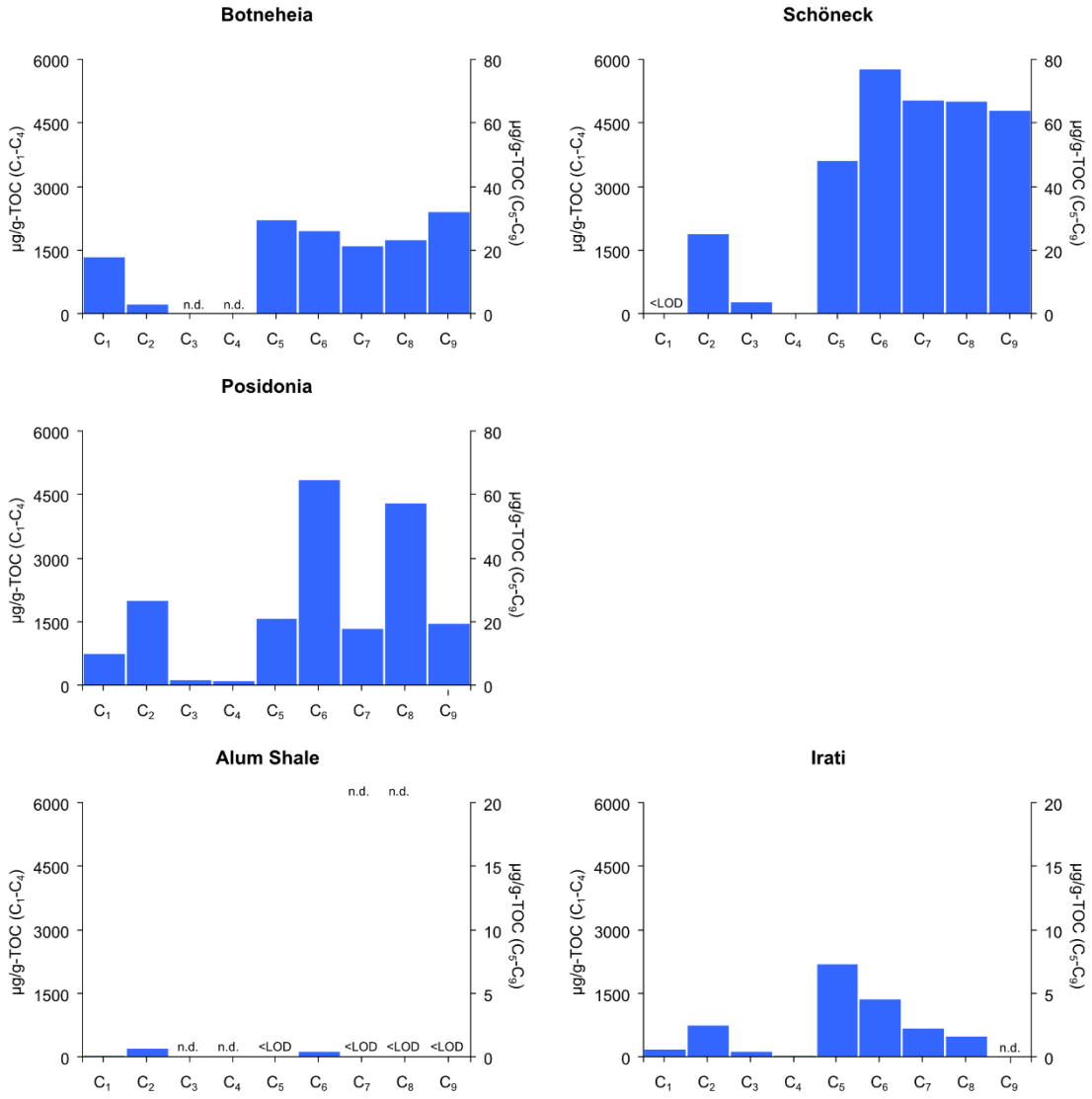


Fig. 6.3 Concentration of saturated carboxylic acids in the water extracts of shales.

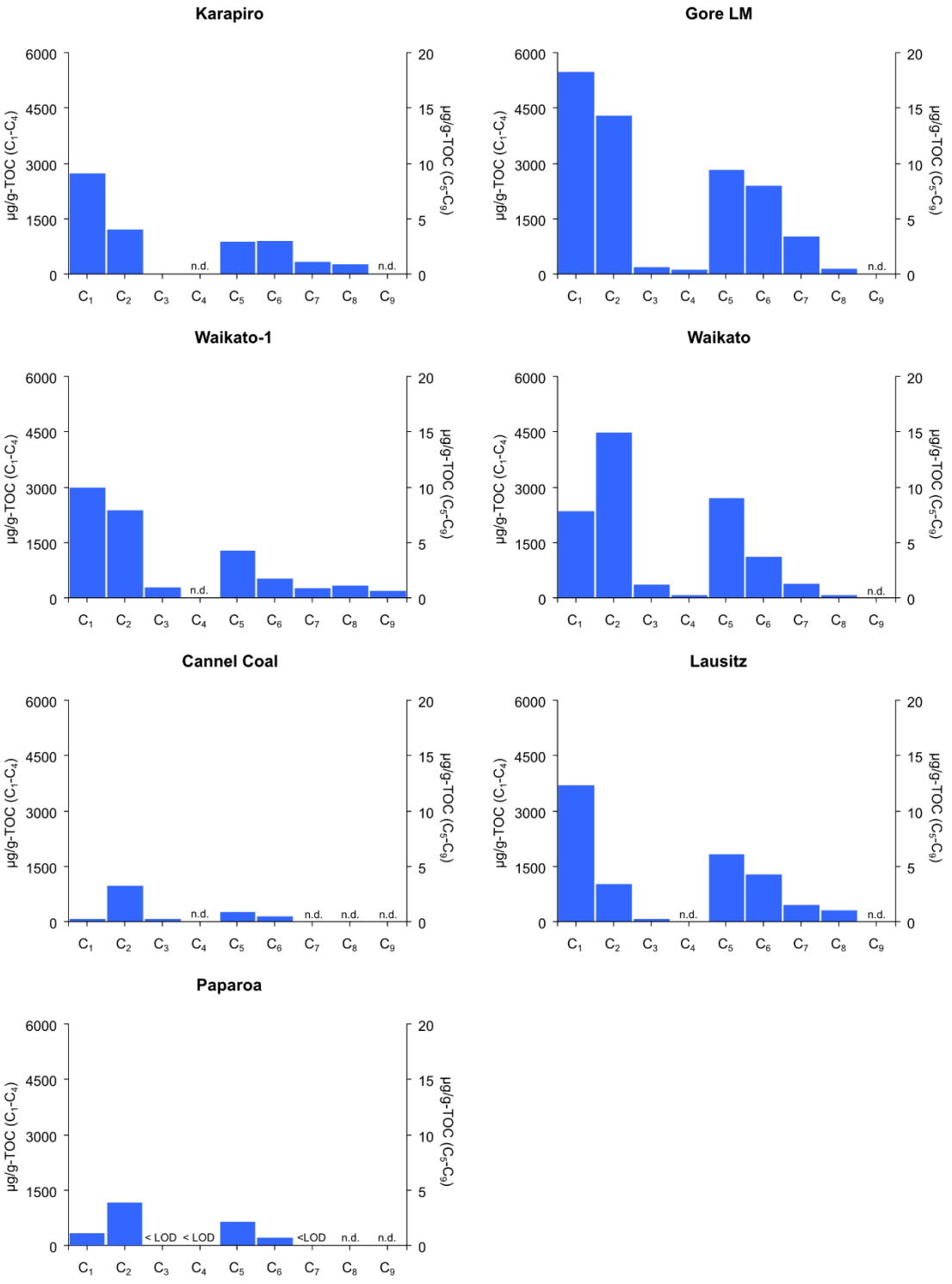


Fig. 6.4 Concentration of saturated carboxylic acids in the water extracts of coals.

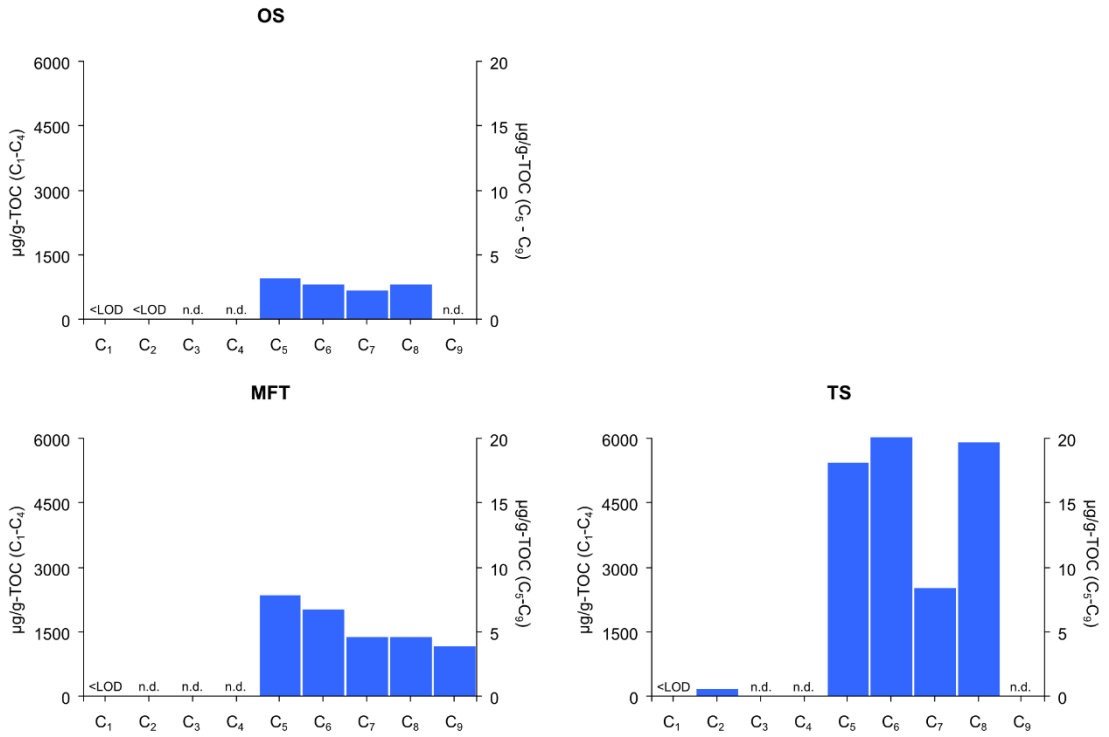


Fig. 6.5 Concentration of saturated carboxylic acids in the water extracts of oil sand-related samples.

In the extracts of oil sand-related samples, the concentration of *n*C5 varied between 3.2-18.1 µg/g TOC, while *n*C6 concentration ranged between 2.7-21.5 µg/g TOC. The content of *n*C7 exhibited values from 2.2 up to 8.4 µg/g TOC. In case of nC8, concentration was between 2.7 and 19.7 µg/g TOC. *n*C9 was only detected in the extract of TS sample and had a concentration of 3.9 µg/g TOC.

Keto and dicarboxylic acids

In the extract of Alum Shale neither keto nor diacids were detected (Fig. 6.6), while in the extracts of Botneheia and Posidonia only oxalic acid was detected, with concentrations of 1494.7 µg/g-TOC and 582.2 µg/g-TOC respectively. Methylsuccinic acid (2255 µg/g-TOC) was detected only in the extract of Schöneck shale. In the extract of Irati shale pyruvic and oxalic acid were only detected, with concentrations of 95.8 and 107.8 µg/g-TOC respectively. Malic acid was not identified in any of the extracts of shales.

In extracts of coals, concentration of oxalic acid ranged from 135.7 µg/g-TOC (Cannel Coal) to 3539.7 µg/g-TOC (Karapiro), whereas pyruvic acid content varied between 27.1 and 1307.1 µg/g-TOC. The malic acid concentrations ranged from 15.7 up 813.6 µg/g-TOC (extract of Gore LM), with the exception of Paparoa extract, where none was detected.

In the extracts of oil sand-related samples, the only detected diacid was oxalic acid. Oxalic acid was present in the extracts of MFT and TS samples at concentrations of 377.8 and 3811.6 µg/g-TOC respectively. This acid may have come from vegetable matter as the former (MFT) was collected from the upper part of the drying cell, where leaves and other plant debris could have been brought in by wind, or even from plant growth. The TS sample is from a reclamation site where plants had already grown.

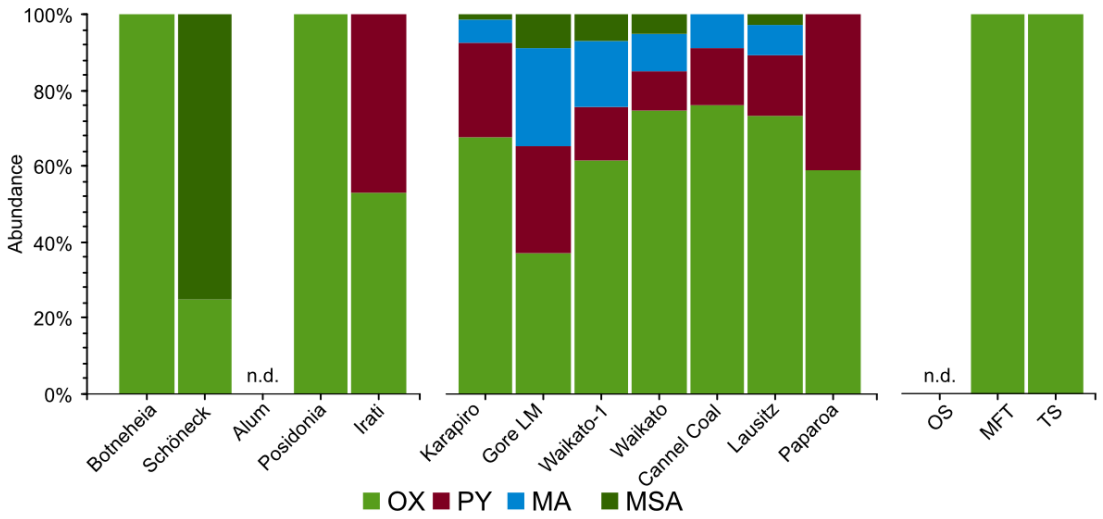


Fig. 6.6 Distribution of water-soluble keto and dicarboxylic acids in the aqueous extracts of the tested samples. OX, oxalic; PY, pyruvic; MA, malic; MSA, methylsuccinic acid.

Inorganic anions

Phosphate was only present in the extracts of Schöneck and Waikato, 12.5 and 26 µg/g sed respectively; while nitrate was only found in extracts of shales (from 5 to 9.8 µg/g sed), but not in the Alum Shale extract. The fluoride content was 10.2 µg/g sed in the extract of Botneheia, the Posidonia extract exhibited a concentration of 11.2 µg/g sed, while that of Irati Fm was 21.4 µg/g sed. The content of this ion was 36.4 µg/g sed in extract of Alum shale. In coal extracts, the concentrations of this ion varied between 7.5 µg/g sed in extract of Waikato1, 8.3 µg/g sed in that of Lausitz, and 10.5, 10.7 and

11 µg/g sed in those of Waikato, Cannel Coal and Paparoa respectively. The fluoride content in the extracts of Gore LM and Karapiro was 13.7 and 21.4 µg/g sed respectively.

The concentration of chloride increased as follows: 9.8 µg/g sed in the extract of Irati, 13.7 µg/g sed in that of Botneheia, 22.1 µg/g sed in extract of Schöneck, and 26.1 and 26.9 µg/g sed in the extracts of Alum and Posidonia shale, respectively. In the extracts of coals, amounts of this ion varied over a wider range. The extracts of Waikato and Waikato-1 exhibited concentrations of 4.4 and 8.6 µg/g sed respectively; the chloride content in extracts of Lausitz and Gore LM were 14.1 and 23.5 µg/g sed, whereas it was 48.1 in the extract of Karapiro; the extract of Paparoa reached 563.4 µg/g sed of chloride. While concentrations of sulphate in extracts of shales were in the similar order of magnitude (thousands) in the shale extracts 4214.1 (Botneheia) to 18843.1 µg/g sed (Posidonia shale), the coal extracts exhibited very different sulphate concentrations, namely 52 and 68 µg/g sed in extracts of Paparoa and Waikato, respectively; 357.9, 558.6 and 801.3 µg/g sed in those of Waikato-1, Lausitz and Karapiro respectively. The sulphate content reached concentrations of 1227.2 and 11256.1 µg/g sed in the extracts of Gore LM and Cannel coal respectively.

The extract of OS exhibited the following concentrations: 1 µg/g sed of fluoride; 20 µg/g sed of chloride and 116 µg/g sed of sulphate; whereas that of TS showed valued of 1.8, 8.3 and 215.4 µg/g sed, respectively. Concentrations of the inorganic ions in the extracts of MFT may reflect the compounds that have been added during the conditioning processes.

These anions are thought to be released from minerals and inorganic phases present in the samples, e.g. fluoride may be formed from fluorapatite and clays, and sulphate from pyrite. However, no identification of the mineral composition was done for these samples.

6.3 Controls on acid release

In order to get an insight into the possible factors that are linked to the release of the acids, the relationships between acid content, kerogen type, geological age, TOC, maturity and inorganic ions were examined.

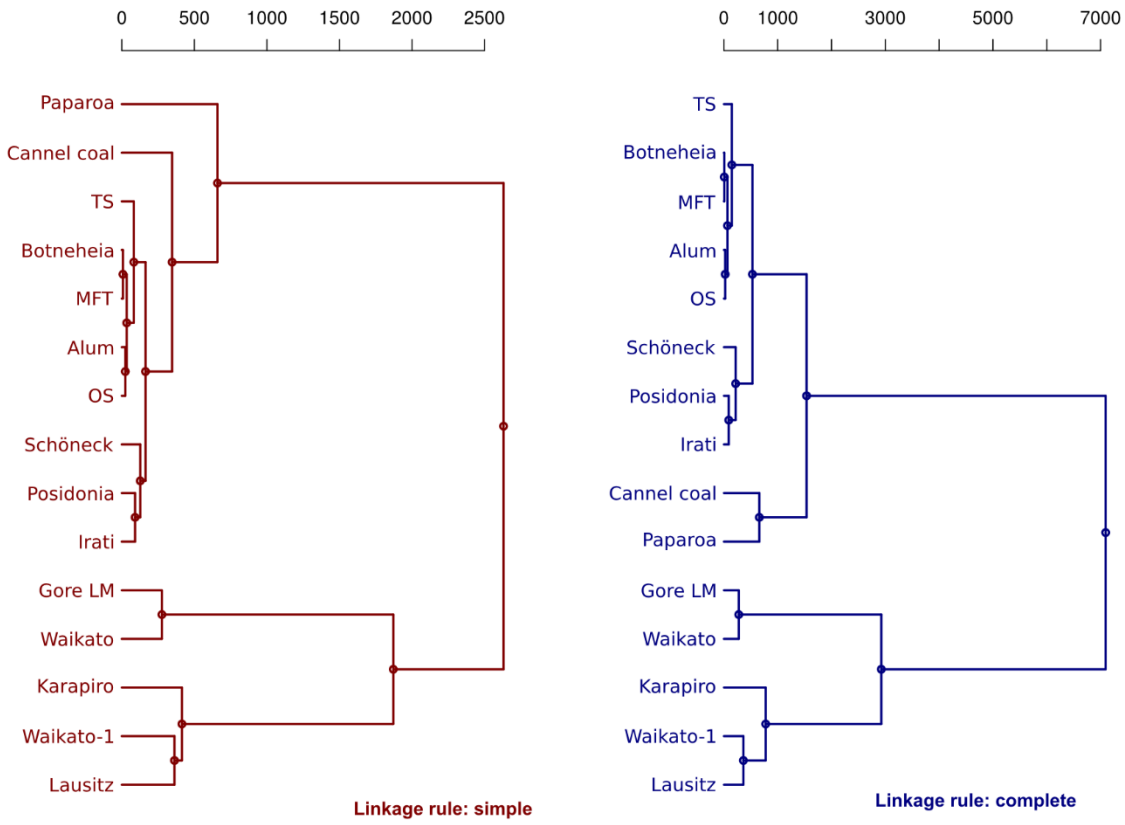


Fig. 6.7 Two dendograms from a hierarchical cluster analysis applied to acid content quantified by IC and expressed in µg/g-sed.

6.3.1 Hierarchical cluster analysis

Hierarchical cluster analysis (Fig. 6.7, Appendix 6) was performed in order to identify any grouping amongst the LMWOA quantified by IC. Single and complete methods were employed as linkage rules. In the former, the distance between two clusters is determined by the distance of the two closest objects (nearest neighbors) in the

different clusters; whereas in the latter, the distances between clusters are determined by the greatest distance between any two objects in the different clusters (i.e., by the "furthest neighbors") (Wessa, 2016).

The dendrogram with single linkage show two main groups. One group comprises 5 coals. These coals are divided in two subgroups: Gore LM and Waikato; and Karapiro, Waikato-1 and Lausitz. The second group includes the five shales, the three oil-sand related samples and two coals, Paparoa and Cannel Coal. These samples were arranged in subgroups as well. Schöneck, Posidonia and Irati shales conform one of the subgroups. The dendrogram with complete linkage also classifies the sample set in two main groups. Again, one of the groups is composed of Gore LM and Waikato; and Karapiro, Waikato-1 and Lausitz coals; and Paparoa and Cannel Coal coals, the five shales and the three oil-related samples constitute the other group. In this heterogeneous subset, Schöneck, Posidonia and Irati shales also form a subgroup. In comparison to the dendrogram with single linkage, this time Cannel Coal and Paparoa coal were grouped together.

Obviously, the division is based on the acid content, thus the extracts of Gore LM and Waikato; and Karapiro, Waikato-1 and Lausitz, which exhibited the largest amount, comprised a subset. In the same way, it is not surprising that the extracts of Schöneck, Posidonia and Irati also formed a subgroup. However, in the dendrogram with a complete linkage, the extracts of Paparoa and Cannel coal, were grouped together. The TOC in the Cannel coal is 67.9 %, while Paparoa's is 74.18 %; whereas the HI ranges between 187 (Cannel coal) and 267 (Paparoa) mgHC/g-TOC. These samples exhibit low OI index (6 and 8 respectively) and similar Tmax (431 and 430 °C respectively). It is interesting that from a cluster analysis of the acid content (quantified by IC), samples were classified according the OI as well.

Hierarchical cluster analysis with complete linkage applied to acid content quantified by IC, could identify a possible relationship between OI and Tmax, and the amount of water-soluble LMWOA that are released from coals and shales.

6.3.2 Factors: Kerogen type, geological age

The extracts of younger sediments, from Paleogene and Quaternary, exhibited concentrations larger than 5000 µg/g-TOC (Fig. 6.8); whereas those of Middle Triassic (Botneheia) and Toarcian (Posidonia) ranged over 3000 µg/g-TOC. The extract of Paparoa coal, late Cretaceous, presented concentration just above 2000 µg/g-TOC, followed by extracts of samples from Permian (Irati) and Carboniferous (Cannel Coal), in which the acid content was around to 1200 µg/g-TOC. The most recent sediments exhibited the largest amounts of acids, between 5155 and 13210 µg/g-TOC. This suggests that to some extent the acid content in the aqueous extracts is related to geological age, i.e. younger sediments may release higher concentrations of acids in comparison to old ones.

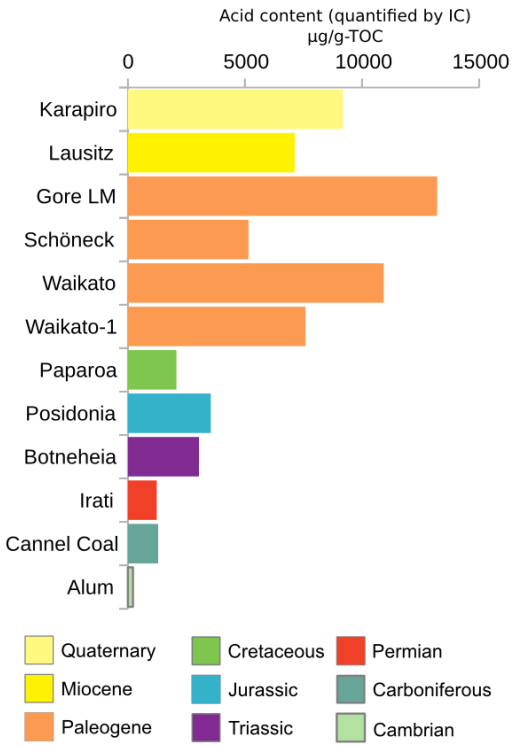


Fig. 6.8 Bar diagram showing acid content (quantified by IC expressed as µg/g-TOC) and geological age.

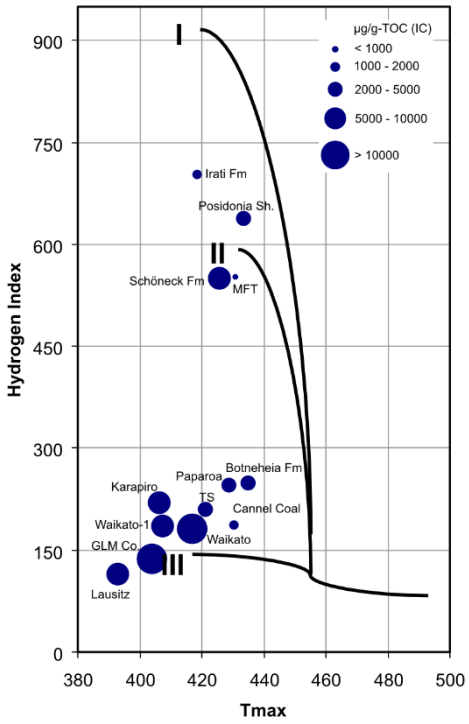


Fig. 6.9 Rock-Eval HI vs Tmax diagram including acid content (quantified by IC).

A modified plot of Rock-Eval parameter hydrogen index and Tmax was employed to evaluate the qualitative relationship between acid content and kerogen type (Fig. 6.9). The data corresponding to Alum shale, the oldest sample from Furogian, is not shown as its Tmax is higher than 500 °C. Amongst the analyzed samples, the sediments corresponding to kerogen type III, which also contain high TOC content, exhibited great amounts of acids. However, Irati shale (kerogen type I and TOC = 25.3 % wt) and Cannel coal (kerogen type I/III and TOC = 68.9 % wt) presented similar acid content.

A Student's *t*–test was used to determine if the acid content averages (expressed as µg/g-TOC) are significantly different for the two categories (Table 6.3): shale and coal extracts. This test was performed using the spreadsheet by John H. McDonald PhD (2014). The results are a t-value of 2.23, 10 degrees of freedom and *P*-value of 0.050 for a two-tailed test, which reject the null hypothesis. A P-value greater than 0.05 (> 0.05) indicates no significant difference detected between the means of the two samples; thus, the amount of acid quantified by IC in the extract of coals is significantly different than that in the extracts of shales. This result suggests that organic matter type also plays a role. Additionally, this dependence will also be discussed based on Pearson's coefficient, discussed lines below.

Table 6.3 Acid content (µg/g-TOC) regarding sample type

Shales	Coals
3038	9189.2
5155.9	13210.9
217.6	7587.2
3541.8	10928.3
1234.2	1293.3
	7128
	2072.5
Average: 2637.5	**7344.2**

6.3.3 Factors: TOC, HI, OI and maturity

Pearson correlation coefficient was calculated to get an insight into the effect of TOC on the acid content released into the water matrix, using both µg/g-sed and µg/g-TOC. This was only applied to the sum of the acids quantified by IC, because the concentration was much higher than that determined by SPE-GC-FID. As explained above, oil-sand related samples relate specific environmental conditions and they were therefore not included in the calculations.

Pearson correlation has been calculated using the whole sample set (Fig. 6.10). It showed that the acid release (µg/g-sed) was positively correlated with OI and TOC, though it was slightly stronger with the former ($r = 0.80$, $R^2 = 0.64$) than with the latter ($r = 0.64$, $R^2 = 0.41$). A moderate linear correlation was observed between acid content Tmax and Calc. Ro, as maturity indicators ($r = -0.48$, $R^2 = 0.23$). However, when the values of the most mature sample (Alum Shale) were removed (plots with empty triangles in Fig. 6.10), a negative linear correlation could be seen ($r = -0.77$, $R^2 = 0.59$). When this coefficient was calculated using IC concentration normalized to TOC (Table 6.4), only OI exhibited a significant positive correlation ($r = 0.86$, $R^2 = 0.74$). This result is similar to that reported by Lundegard and Senftle, (1987). While testing the yield of acid under different temperatures of hydropyrolysis, they observed that the sample with highest OI but lowest TOC, yielded the largest acid content at all tested temperatures.

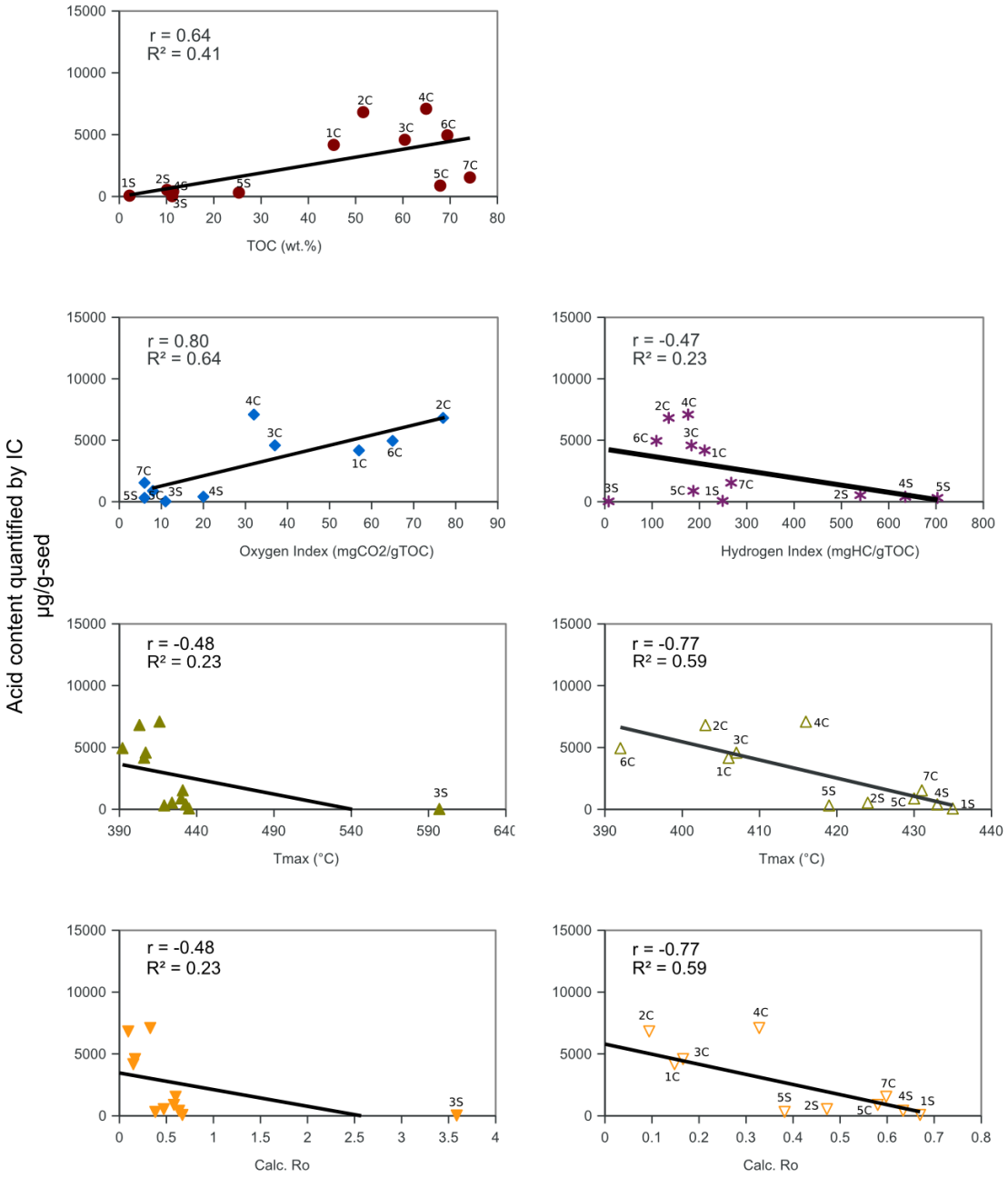


Fig. 6.10 Plots showing the correlation between the content of water-extractable acids and some sample properties. Calc. Ro according Jarvie et al., (2001). Empty triangles refer a subsample set in which the data of Alum shale extract were not included. See Table 6.1 for nomenclature and text for explanation. r, Pearson correlation coefficient; R^2, coefficient of determination.

Table 6.4 Pearson correlation coefficient of acid content (µg/g-TOC) with some sample properties

Parameter	r	R^2	r*	R^{2*}
OI	0.86	0.74		
HI	-0.30	0.09		
Tmax	-0.54	0.29	-0.77	0.49
Calc. Ro	-0.54	0.29	-0.77	0.49

*Data of Alum shale extract were not included.

Linear correlation was also calculated for Tmax and calc. Ro based on the sample type. The data of Alum Shale were not included to avoid masking the correlation as explained previously. Expressing concentration as µg/g-sed (Table 6.5), the acid content in the extracts of shales showed a strong positive correlation with HI (r = 0.82, R^2 = 0.67), whereas negative correlation was observed for maturity (Tmax and Calc. Ro); weak correlation of acid content was observed with TOC and OI, r = 0.30 and R^2 = 0.09, and r = 0.39 and R^2 = 0.15, respectively. Regarding the coal extracts, the strongest correlation of acid content was observed with OI (r = 0.73, R^2 = 0.54), whereas it was also also strong but negative with TOC, Tmax, HI and Sr.

Table 6.5 Pearson correlation coefficient of acid content (µg/g-sed) regarding sample type

	Shales		Coals	
Parameter	**r**	**R^2**	**r**	**R^2**
TOC (wt%)	0.30	0.09	-0.68	0.46
OI	0.39	0.15	0.73	0.54
HI	0.82	0.67	-0.6	-0.36
Tmax*	-0.48	0.23	-0.68	0.46
Calc. Ro*	-0.61	0.38		
Sr			-0.51	0.26

*Data of Alum shale extract were not included.

The variations in the coefficient of determination regarding sample type if compared with the calculations using the whole set of samples, are assumed to be due to the differences between samples become stronger, as sample size is reduced. For instance, the correlation between acid content and HI was stronger in shales, while the correlation with OI was stronger in the coal extracts. This can be related to the fact that HI tends to be higher in shales than in coals, as OI is higher in coals than in shales. Such results accord with the Student's *t*–test previously calculated, thus, the organic matter type to some extent affects the release of LMWOA.

Calculations using µg/g-TOC (Table 6.6), show that in the extracts of shales the correlation between acid content with OI continued being positively strong; while Pearson coefficient behaved moderately positive in case of HI. The maturity indicators (Tmax and Calc. Ro) exhibited a weak correlation with the amount of acids. In the coal extracts, the acid content is significantly correlated with OI. Suggate Rank also showed strong negative correlation with the acid concentration, and Pearson's indicator showed a negative correlation (r = -0.68) in case of Tmax.

Table 6.6 Pearson correlation coefficient of acid content (µg/g-TOC) regarding sample type

	Shales		Coals	
Parameter	**r**	**R^2**	**r**	**R^2**
OI	0.79	0.62	0.81	0.66
HI	0.45	0.21	-0.53	0.28
Tmax*	0.29	0.08	-0.68	0.46
Ro*	-0.69	0.47		
Sr			-0.74	0.55

*Data of Alum shale extract were not included.

Summarizing, this analysis suggests that OI has a larger effect on the amount of released water-extractable acid, followed by maturity; and as it is to some extent the organic matter type. This accords with the fact that immature organic matter comprises a great amount of numerous oxygenated functional groups, which are lost during late diagenesis and catagenesis (Hetényi, 1998; Vu, 2008).

6.4 Multiple regression applied to LMWOA released from coals

Once concentrations have been determined, a model to estimate the amount of LMWOA in new samples is desirable. Multiple regression (MR) is usually used to find an equation that best predicts the dependent variable as a linear function of the independent variables. In order to prevent poor predictive performance, as rule of thumb, MR requires between 10 and 20 times as many observations as independent variables (McDonald, 2014). As previously discussed, any estimation of LWMOA with the IC procedure used in this study, should be done according to the organic matter type and should at least involve OI and maturity as independent variables. The two datasets used in this study consist of 5 shales and 7 coals, which means a sample size smaller than the one required to apply MR. However, MR could be used as preliminary test to assess whether a formula with two variables is significantly better than a formula with just one predictive variable. This procedure was applied to the coal subset, as there are just a few OI data for the shales.

The null hypothesis of MR states that there is no relationship between OI and Tmax, with the acid content quantified by IC in aqueous extracts of coals, i.e. the sum of formic, acetic, propanoic, butanoic, oxalic, piruvic, malic and methylsuccinic acid. Tmax was employed as maturity parameter because MLS requirement of none missing

values in the data set. The MR was performed with a spreadsheet by McDonald (2014), expressing concentration as µg/g-sed. The calculations showed that the optimized equation involves only OI as independent variable with a R^2 of 0.54 and a P-Value of 0.06 (Eq. 1, Table 6.7 and Appendix 5).

Table 6.7 Results of multiple regression applied to acid content quantified by IC (µg/g-sed)

# Eq.	Number of observations	Sample not included	R^2	P-value of R^2	Regression equation
1	7	--	**0.54**	**0.06**	**Y = 1737.1713 + 63.35815*OI**
2	6	Karapiro	0.58	0.08	Y = 1728.8581 + 68.82244*OI
3	6	GLM	0.40	0.17	Y = 1803.3959 + 60.43280*OI
4	6	Waikato-1	0.54	0.09	Y = 1636.1359 + 63.78360*OI
5	6	Waikato	**0.89**	**0.004**	**Y = 881.13598 + 70.58313*OI**
6	6	Cannel Coal	0.37	0.2	Y = 2672.7376 + 47.85683*OI
7	6	Lausitz	0.56	0.09	Y = 1640.7375 + 70.21232*OI
8	6	Paparoa	0.40	0.18	Y = 2171.0729 + 56.02667*OI

For interpretation, R^2 values closer to unit and small P-Values indicate the best fit. In order to get an insight into the possible effect of each coal on the equations, MR was also calculated excluding one coal each time (Table 6.7). R^2 ranged between 0.37 up to 0.89, whereas the P-values varied between 0.004 and 0.2. R2. The equation that exhibited the best fit, was the one not including the Waikato coal, R^2 = 0.89 and P-Value = 0.004 (Eq. 5, Table 6.7). Amongst these coals, Waikato-1 is most similar to Waikato: similar TOC, OI and age. The only difference between these two samples is related to the organic matter. Vu (2008) reported that the organic matter in Waikato contained planktonic algal components. This result may support the previous discussion about how the organic matter also affects the release of LMWOA.

The influence of OI on the release of LMWOA was demonstrated, thus a classification of the sediments in regard of the released acid content can be done based on this parameter. Hierarchical clustering can serve to illustrate it (Fig. 6.11). Based on the shales and coals in this study, sediments with OI between 8 and 20 mgCO2/g-TOC would release less amounts of acids in comparison to sediments with OI between 32 and 77 mgCO2/g-TOC.

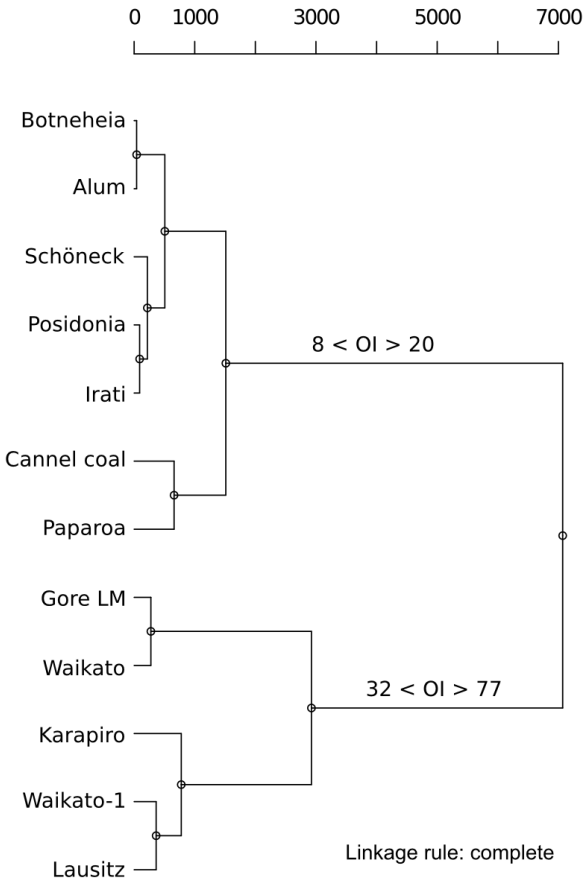


Fig. 6.11 Dendrogram from a hierarchical cluster analysis applied to acid content (quantified by IC expressed in µg/g-sed) in the extracts of shales and coals.

6.5 Relationship of inorganic ions with selected carboxylic acids

Pearson correlation coefficient (using both µg/g-TOC and µg/g sed as units) between fluoride, chloride and sulfate and most abundant carboxylic acids was chosen to understand the mechanism by which the extraction of inorganic ions may affect the release of carboxylic acids in the extracts of shales and coals. In the extracts of shales, a moderate negative correlation (r = -0.71, R2 = 0.51) was observed between formic acid and fluoride; whereas this acid only weakly negatively correlated with chloride (Table 6.8), regardless of which concentration unit was used. In extracts of shales, the weak correlation of acetic acid with fluoride chloride became stronger when µg/g-TOC was used. However, a change in concentration unit did not affect the correlation between sulfate and acetic acid. Weakly negative correlations were observed between acetic acid and the selected inorganic anions in the extracts of coals (Table 6.9). In the

extracts of shales, oxalic acid exhibited a strong negative correlation with fluoride using µg/g-TOC as concentration, whereas this acid showed strong positive correlation with chloride when correlation was undertaken using µg/g-sed. In the extracts of coals, correlation between fluoride and oxalic acid become stronger by using the concentration normalized to TOC (Table 6.10).

Table 6.8 Pearson correlation coefficient of formic acid and selected inorganic anions

	Shales		Coals	
	r (R^2)	r (R^2)	r (R^2)	r (R^2)
Parameter	µg/g-sed	µg/g-TOC	µg/g-sed	µg/g-TOC
F^-	-0.71 (0.51)	-0.65 (0.43)	-0.09 (0.01)	0.12 (0.01)
Cl^-	-0.17 (0.03)	-0.23 (0.06)	-0.58 (0.33)	-0.57 (0.32)
$SO4^{2-}$	0.28 (0.08)	-0.31 (0.09)	-0.53 (0.28)	-0.5 (0.25)

Table 6.9 Pearson correlation coefficient of acetic acid and selected inorganic anions

	Shales		Coals	
	r (R^2)	r (R^2)	r (R^2)	r (R^2)
Parameter	µg/g-sed	µg/g-TOC	µg/g-sed	µg/g-TOC
F-	-0.37 (0.14)	-0.51 (0.26)	-0.20 (0.04)	-0.07
Cl-	0.11 (0.01)	0.42 (0.18)	-0.3 (0.09)	-0.36 (0.13)
SO42-	0.24 (0.06)	0.25 (0.07)	-0.33 (0.11)	-0.33 (0.11)

Table 6.10 Pearson correlation coefficient of oxalic acid and selected inorganic anions

	Shales		Coals	
	r (R^2)	r (R^2)	r (R^2)	r (R^2)
Parameter	µg/g-sed	µg/g-TOC	µg/g-sed	µg/g-TOC
F-	-0.33 (0.11)	-0.81 (0.65)	0.34 (0.12)	0.59 (0.35)
Cl-	0.91 (0.83)	0.05 (0.002)	-0.48 (0.23)	-0.46 (0.21)
SO42-	0.47 (0.22)	-0.48 (0.23)	-0.54 (0.29)	-0.49 (0.24)

Based on these results, it can be concluded that oxalic acid was more strongly related to inorganic ions, specifically fluoride and chloride. The inverse linear correlation may be explained by the possible sorption of fluoride in the sediment and/or formation of insoluble fluoride salts (Pickering, 1985), where more oxalic acid molecules may be released into the aqueous matrix. This observed effect on the diacid could be because oxalic acid is more reactive than formic and acetic acid. Recently Wilke et al. (2015) published concentrations of several cations and some carboxylic acids released from black shales extracted with water or mixtures of electrolytes, during 24 h or up to 6

months. However, they only reported that fluoride was detected in the water extracts of few samples and in low concentrations using an extraction time of 24 h. A deeper characterization of fluoride content in shales and coals, before and after extraction, would be necessary to clarify the processes controlling the apparent relationship between fluoride and oxalic acid in the extracts.

6.6 Pattern distribution of saturated LMWOA

Figs. 6.3-6.5 show that the concentration of saturated carboxylic acids in extracts of shales and coals exhibit different distribution regarding carbon chain length. The extracts of shales did not show a specific pattern. For example, the extract of Schöneck shale did not clearly exhibit a preferential odd or even-carbon chain, whereas that of Posidonia did for an even-carbon chain length. Moreover, acetic acid was predominant in most of these extracts. However, in most extracts of coals, concentration of formic acid was higher than that of acetic acid, and a decrease of concentration could be seen with an increase in the carbon number in the chain, namely *n*C5 concentration was higher than *n*C6, which in turn exhibited a higher concentration than *n*C7 and so on.

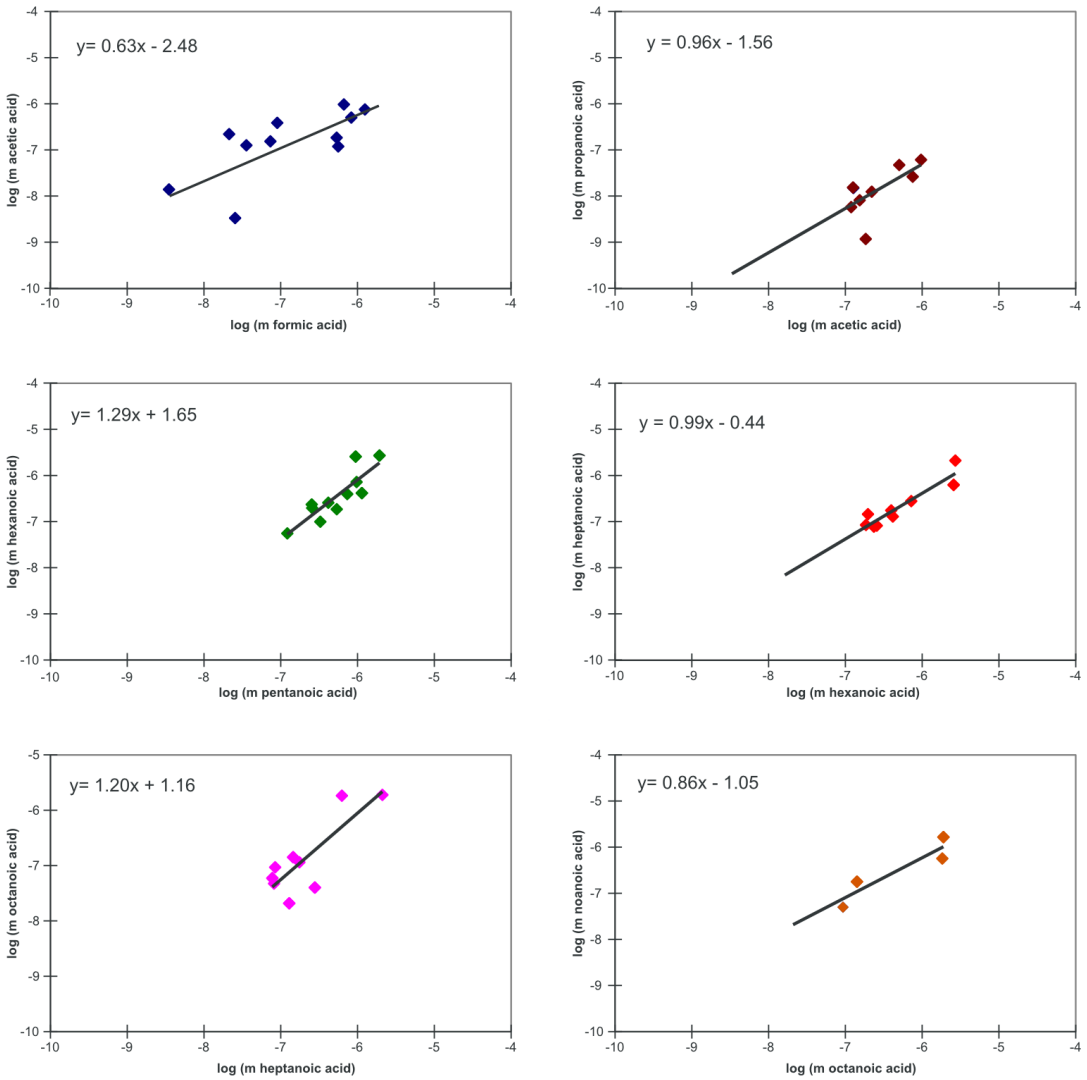


Fig. 6.12 Comparison of abundances of saturated acids, C1-C2, C2-C3, C5-C6, C6-C7, C7-C8 and C8-C9, in water extracts of shales and coals.

Acetic acid is a major carboxylic acid found in natural waters because it is a key intermediate in the oxidative degradation of DOC by bacteria (Thurman, 1985), but it could also be produced by means of inorganic reactions (Seewald, 2001). The latter reports that the thermal oxidation of alkanes results in the formation of alkenes, alcohols, ketones and carboxylic acids (final product). In his paper, Seewald uses laboratory results of short-chain organic acids produced by oxidation of C2-C6 saturated alkanes to provide a model that predicts the amount of carboxylic acid that would be produced from larger carbon-chain alkanes. To support the model, Seewald employed log vs log (molar concentration) plots showing the abundance of formic in relation to acetic acid, and again acetic acid in relation to propanoic acid in basinal

brines. The regression line should exhibit a slope close to 1, to confirm the origin of carboxylic acids via oxidative degradation of alkanes. In the present study, to explore the results by Seewald, similar plots were prepared using the concentrations of acids with carbon-chain length from C1 up to C9 in the water extracts of coal and shales (Fig. 6.12). The data of oil sand-related samples were not included. It is worth mentioning that molarity was used as concentration unit in these calculations as the study by Seewald, in order to reduce the disparity in molecular weight amongst acids that can affect the comparison.

As it is shown in Fig 6.12, the slopes of regression lines showed values between 0.63 and 1.29, i.e. are not exactly equal to unity. However, Seewald's model may still explain the distribution of the concentration of carboxylic acids in the aqueous extracts. The deviations in the slopes may be related to biotic process affecting the production of these compounds. For example, in the extracts of Schöneck and Posidonia shale (both from marine depositional environments), concentrations of carboxylic acids with carbon number between 5 and 9, are much higher than those in the extracts of coals. No decrease in concentration was observed with an increase in the carbon chain in these extracts, as it was observed in the most extracts of coals.

Regarding oil sand-related samples, (Fig. 6.5), only acetic acid was detected in the extract of TS and although its concentration was lower in comparison with other samples, it may be indicative of both the ongoing microbial activity in the reclamation site and the contribution of the added peat, especially considering the concentrations of *n*C5-*n*C9 carboxylic acids, which were higher than those found in the extracts of coals. Interestingly, in TS the pattern of the concentration of these acids is similar to that of the shales. It is not surprising that the extracts of OS and MFT also exhibited carboxylic acids with chain length between 5 and 9 carbons. First, because oil sands are the product of the microbial alteration and secondly Noah *et al.* (2014) suggested that mature fine tailings of Alberta basin have also undergone biodegradation to some extent.

6.7 Changing extraction conditions

To know to what extent the yields of water-soluble LMWOA may be affected, different extraction conditions were tested on Waikato-1 coal and Posidonia shale. These samples were tested under two conditions: (i) three different water/sediment ratios, by varying the amount of sediment, specifically 0.5 g for a 500:1 ratio; 2.5 g for 100:1 ratio and 5 g corresponding to 50:1 ratio; and, (ii) pH and phosphate content, the extractions were done with 5 g of sediments in aqueous buffer solutions of pH 2, pH 7, and 12.5. All extractions were carried out in a fixed amount of Millipore water of 250 ml, at temperature of 80 - 90 °C during 72 h; more details about extraction procedure can be found in Chapter 3.

6.7.1 Effect of water/sediment ratio

After extraction, the pH in the extracts were as follows: for Waikato-1 coal: pH 4.7 for 500:1 ratio; pH 5.6 for 100:1 ratio and pH 4.3 for 50:1 ratio. The extracts of Posidonia exhibited pH 6.3 for 500:1 ratio; pH 7.2 for 100:1 ratio, and pH 7.4 for 50:1 ratio. Formic acid in Waikato-1 water extracts ranged from 5.4 (500:1 ratio) to 38.2 mg/l (50:1 ratio, Fig. 6.13), whereas the values were between 0.6 (500:1) and 2.7 mg/l (50:1) in the water extracts of Posidonia shale. The concentrations of acetic acid varied from 4.3 (500:1) to 30.2 mg/l (50:1) in Waikato extracts; while those of Posidonia shale exhibited concentrations of this acid between 0.7 (500:1) and 6 mg/l (50:1). In the extracts of Waikato-1, amounts of propionic acid fluctuated from 0.4 (500:1) to 3.5 mg/l (50:1), whereas in those of Posidonia, concentrations were between 0.1 (100:1) and 0.6 mg/l (50:1), in the extract of 500:1 ratio, this acid was not detected.

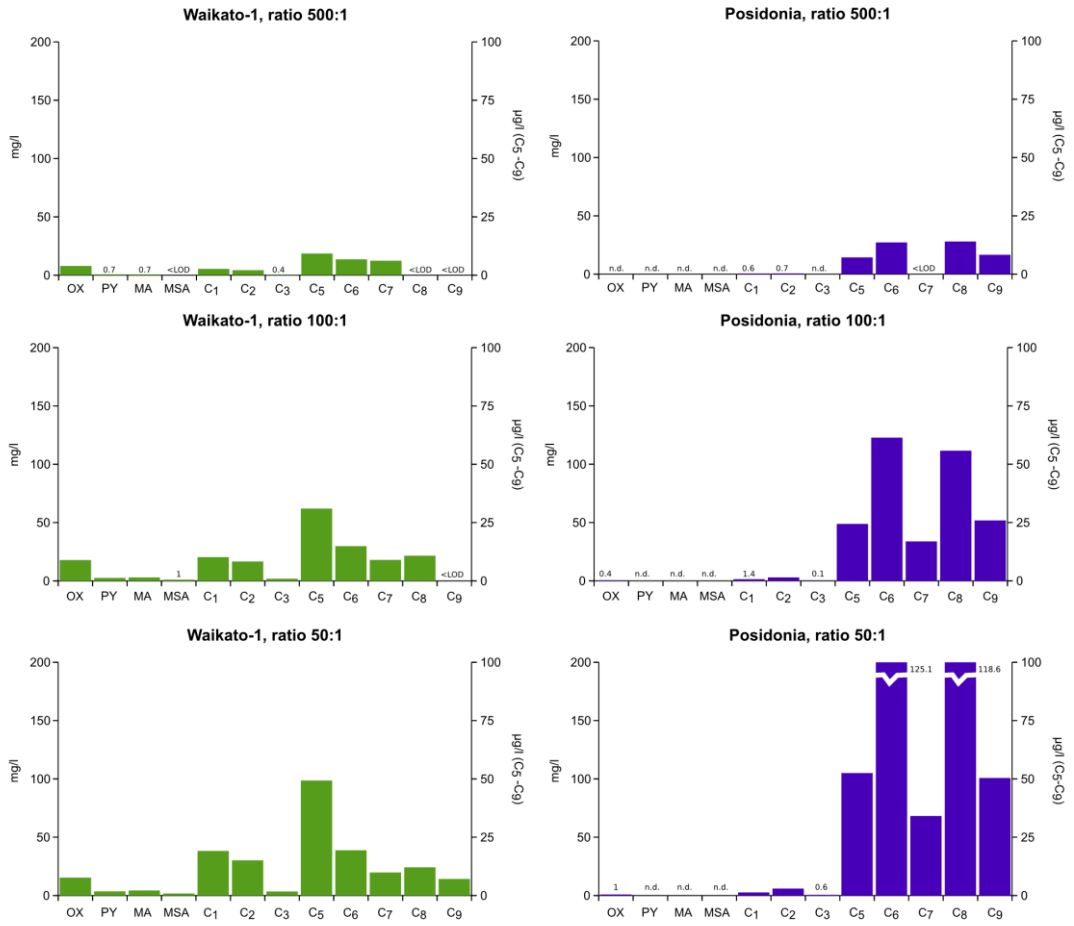


Fig. 6.13 Concentrations of carboxylic acids, expressed as mg/l (or µg/l), in the water extracts of Waikato-1 coal (green bars) and Posidonia shale (blue bars) under different water/sediment ratios. OX, oxalic; PY, pyruvic; MA, malic; MSA, methylsuccinic acid; Cn, *n*-saturated acid by carbon number.

In the extracts of Waikato-1, concentrations of pentanoic acid varied from 9.3 (500:1) to 49.3 µg/l (50:1), but between 7.2 (500:1) and 52.5 µg/l (50:1) in the extracts of Posidonia shale. Concentrations of hexanoic acid ranged from 6.8 (500:1) to 19.4 µg/l (50:1) in extracts of Waikato-1, and between 13.6 (500:1) and 125.1 µg/l (50:1) in those of Posidonia shale. In the extracts of Waikato-1 concentrations of heptanoic were as follows 6.2, 9 and 9.9 mg/l for 500:1, 100:1 and 50:1 ratio respectively; whereas in extracts of Posidonia shale, concentration of this acid was below the detection limit in the extract corresponding to a ratio of 500:1, but was detectable at other ratios varying between 16.9 (100:1) and 34.1 µg/l (50:1). In the extract of Waikato-1, octanoic acid was not detected in the extract using the 500:1 ratio, concentrations were 10.8 and 12.1 µg/l for 100:1 and 50:1 ratio respectively. However in the extracts of Posidonia shale, this acid exhibited the following concentrations: 14 (500:1), 55.8 (100:1) and 118.6 µg/l (50:1). Nonanoic acid was detected only in the extract of Waikato-1 in a

water/ sediment ratio of 50:1; while the extracts of Posidonia shale exhibited concentrations ranging from 7.1 (500:1) to 25.9 µg/l (50:1). Thus, in the water extracts of both samples, the concentration of the selected carboxylic acids expressed as mg/l (or µg/l) rises as the amount of sediment employed in the extraction increases.

Fig. 6.14 shows the concentration of released carboxylic acids normalized to TOC content (µg/g-TOC). In extracts of Waikato-1, from lowest (50:1) to highest ratio (500:1), concentrations of formic acid were between 2991.5 and 4385.3 µg/g-TOC; acetic acid ranged from 2364.4 to 3503.2 µg/g-TOC and propanoic acid from 273 to 288.9 µg/g-TOC. Concentrations of oxalic acid varied from 1200 to 6443 µg/g-TOC; pyruvic acid from 283 up to 564 µg/g-TOC; malic acid from 338 to 547 µg/g-TOC, and methylsuccinic acid between 136 and 167 µg/g-TOC. Although all these acids exhibited an increase in yields with the highest ratio (500:1), oxalic acid showed the highest released amount. Concentrations of pentanoic acid ranged from 8.4 to 4.3 µg/g-TOC; hexanoic acid from 6.1 to 1.7 µg/g-TOC; heptanoic acid from 5.6 to 0.9 µg/g-TOC. The concentration of octanoic acid was below of the detection limit in a ratio of 500:1; it could be detected at low values at other ratios (1.9 at 100:1 and 1.1 µg/g-TOC at 50:1). Nonanoic acid could only be quantified in the 50:1 ratio, 0.6 µg/g-TOC.

The extracts of Posidonia shale exhibited concentrations of formic acid between 1098.6 (50:1) and 2395.5 µg/g-TOC (500:1), whereas concentrations of acetic acid varied from 2776.9 (500:1) to 2460 µg/g-TOC (50:1). Propanoic acid was present in concentrations between 93.1 (100:1) and 232.4 µg/g-TOC (50:1), however this acid was not detected in the extract of 500:1 ratio. Concentrations of oxalic acid ranged between 333 (100:1) and 374.4 µg/g-TOC (50:1), but were undetectable in the extract of 500:1 ratio. Pyruvic, malic and methylsuccinic acids in the extracts of this sample. Concentrations of pentanoic acid found in the Posidonia shale extracts varied from 34.4 (500:1) to 24 µg/g-TOC (50:1), those of hexanoic acid between 65 (500:1) to 57.3 µg/g-TOC (50:1); octanoic acid ranged from 66.9 (500:1) to 54.3 µg/g-TOC (50:1), whereas concentrations of nonanoic acid were between 39.7 (500:1) to 23.1 µg/g-TOC (50:1). Heptanoic acid showed the lowest concentrations with 16 and 15.6 µg/g-TOC for 100:1 and 50:1 ratios, respectively.

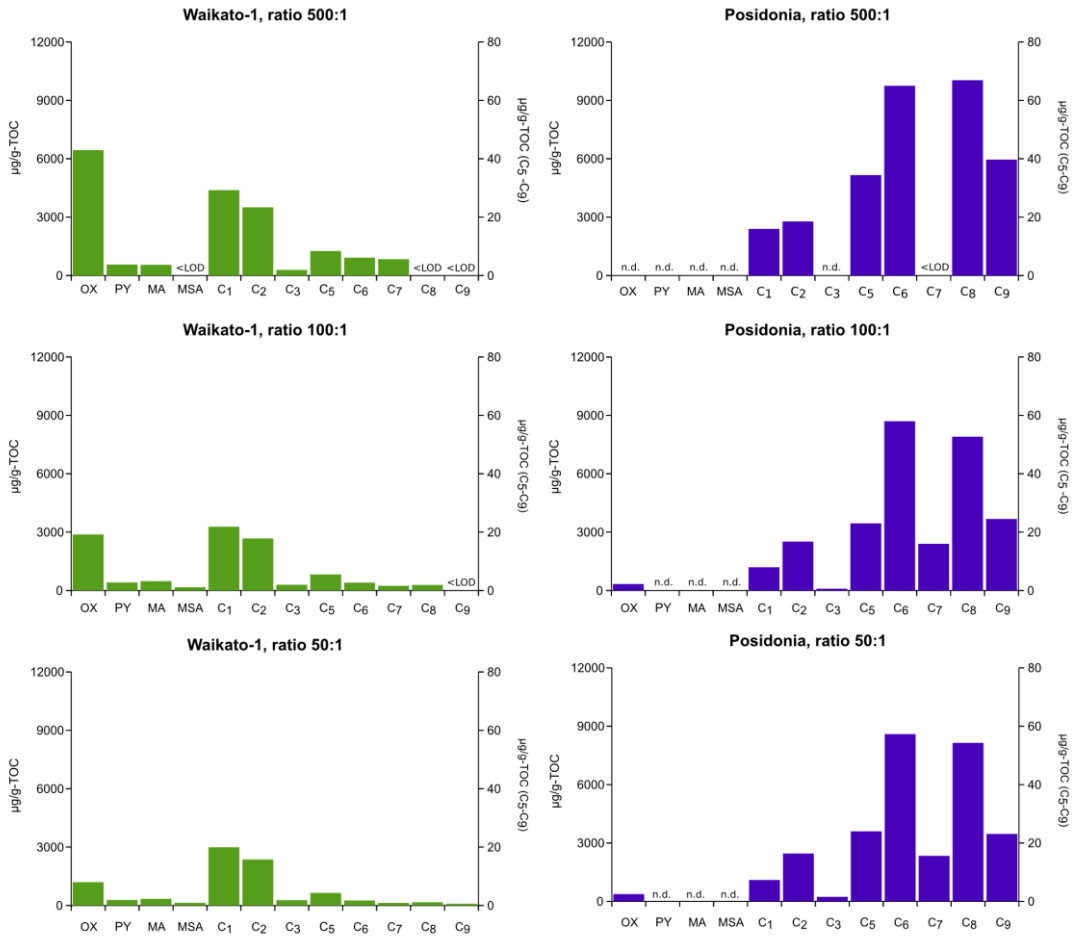


Fig. 6.14 Amounts of selected carboxylic acids, expressed as µg/g-TOC released using different water/sediment ratios. For nomenclature see Fig. 6.13.

Based on these results, it can be observed that in the Waikato-1 extracts, the highest ratio (500:1) yielded the highest amount of carboxylic acids per gram of TOC, whereas the extracts of Posidonia shale did not exhibit such strong variations. This situation seems to be opposite to the results shown in Fig. 6.13 showing plots expressing concentrations as mg/l.

While solubility of the acids does not change during extraction, the intra and intermolecular interactions between the water and compounds within the sediment particles can vary. This is probably more evident in the extracts of coal (Waikato-1) as coals generally possess more diverse composition of oxygen-containing functional groups which may provide a more intricate organic matter structure, thus more interaction sites, in comparison to the shale (Killops and Killops, 2004). Thus, in higher water/sediment ratios, water molecules (of smaller size) could move more efficiently (a

deeper contact) across the structure of the sediment particles; the water excess may have also reduced the re-sorption onto the matrix that some compounds could undergo. Although, it has been long examined, the ions behavior in solution is still not fully understood, thus the ongoing complex solvation of the wide classes of compounds within the sediments during the extraction, should be considered. Solvation (or hydration for aqueous solutions) refers to the structure that surrounds each molecule or ion in solution, which is formed by solvent molecules. This structure is the result of intermolecular forces between solute and solvent. The solvation ability of a particular solvent is the result of mainly four components: i) the energy that the dissolved molecule or ion requires to produce “a hole” in the solvent; ii) the orientation energy of the dipolar solvent molecules produced by the presence of the solvated molecule or ion; iii) the isotropic interaction energy due to intermolecular forces as electrostatic, polarization, and dispersion energy; iv) the energy coming from the formation of hydrogen bonds or electron-pair donor/electron-pair acceptor bonds (Reichardt and Welton, 2011). The water itself exhibits an intricate molecular network, which explains its remarkable properties, for example the formation of ice at low temperatures (Stillinger, 1978). Carboxylic acids also exhibit interesting reactions. For example, acetic acid can form a cyclic dimer by two equivalent hydrogen bonds. This dimer is stable that it can remain in the gas phase. Currently, research is being made to understand the mechanisms of dimer – monomer – ionization of carboxylic acids in water (Yamabe and Tsuchida, 2003; Hermida-Ramón *et al.*, 2004).

Though all acids exhibited an increase in yield with the highest water/sediment ratio (500:1), the increase observed in oxalic acid was remarkable. This may be related to a) adsorption of monocarboxylic acids (e.g. acetic acid) on sediments is weaker than the one of dicarboxylic acids, as oxalic acid (Jones *et al.*, 2003), thus higher amount of water favored the desorption of more oxalic acid; b) the original abundance of this diprotic acid within the complex structure of the coal was higher than other carboxylic acids, which in turn could have been reached by the water excess; c) formation of oxalate by oxidative cleavage at susceptible functional groups for instance 1,2-dihydroxycarboxylic, 1-hydroxy-2-keto- or 1,2-diketocarboxylic acid (Bou-Raad *et al.*, 2000); or, d) the complex solvation mechanism that this compound may undergo. There are at least five known conformational isomers (or conformers) for oxalic acid, which could form several cluster conformations regarding the number of water molecules influencing the inter- and intramolecular hydrogen bonds formed by this acid

(Blair and Thakkar, 2010; Hermida-Ramón *et al.*, 2004). Thus, each carboxylic acid was affected to different extent by the amount of water in relation to amount of sediment. This test was duplicated and confirmed the strong effect that water excess seems to have on the release of oxalic acid. The results also confirm the necessity for methodological standardization of quantification methods for production rates and finally, for data comparison.

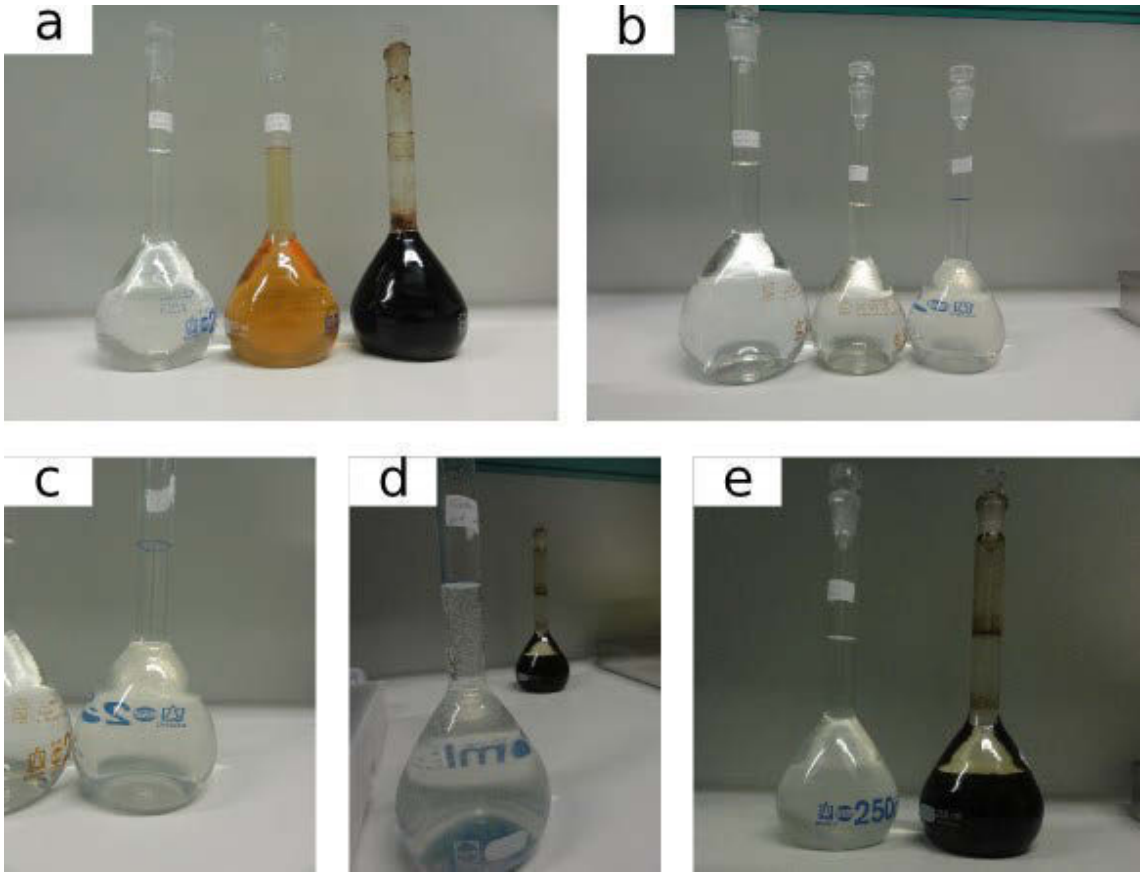


Fig. 6.15 Water extracts of Waikato-1 coal (a) and Posidonia shale (b) using buffered aqueous solutions at pH 2, 7 and 12.5 (ordered from left to right). Formation of an emulsion in the Posidonia extract at pH 12.5 (c) and effervescence (d and e), and precipitation of brownish compounds in the Waikato-1 extract (e, to the right) during acidification for SPE procedure.

6.7.2 Effect of phosphate buffers at different pH

Three (2, 7 and 12.5) initial pH of the water were tested. However, the measured pH in extracts varied as follows 2, 6.8 and 11.1 in extracts of Waikato-1; whereas those in extracts of Posidonia shale were 3.8, 7.7 and 11.7.

In case of Waikato-1 coal, the effect of pH-phosphate during extraction was visibly

evident. The water extract at pH 12.5 exhibited the darkest (Fig. 6.15a) color. The color is related to humic-like substances, e.g. chromophores, phenolic quinones presenting conjugated double bonds. Thus, pH above 7, in which fulvic and humic acids are soluble, may favor the release of such compounds. Subsequently, when the aqueous extract was acidified for SPE procedure, precipitation of a large amount of dissolved substances was observed (Fig. 6.15e to the right). These precipitated compounds are thought to be humic acids which are insoluble at low pH. Thurman (1985) reported that color related to carbon is very similar between soil extracted fulvic acids and aquatic fulvic substances derived from terrestrial matter; in this way the color of water extracts of Waikato-1 at pH > 7 may be related to the terrestrial origin of the organic matter of the coal.

The aqueous extracts of Posidonia shale did not exhibit any change of color among the three tested solutions (Fig. 6.15b). However, when the water extracts were acidified in accordance with the SPE procedure, the most basic extract exhibited the formation of an emulsion (Fig. 6.15c) and effervescence (Fig. 6.15d and e to the left). Although neither carbonate nor metal content in the samples was determined, the Posidonia shale has been reported to contain pyrite (Littke *et al.*, 1991); and calcite (Bernard *et al.*, 2012). This calcite, mainly in form of coccoliths and other plankton-derived microfossils, may have reacted with dipotassium phosphate, favoring the mobilization of metals as Fe and Al and the formation of diverse phosphate compounds as $CaHPO_4 \bullet 2H_2O$ and/or $MgHPO_4 \bullet 3H_2O$ (Lindsay *et al.*, 1961), which may be manifested as the emulsion.

In the extracts of Waikato-1 (Fig. 6.16), concentration of formic acid increased from 25.4 mg/l at pH 2, to 136 mg/l at pH 12.5[3], similarly acetic acid also increased from 34.1 mg/l (pH 2) to 110.2 mg/l (pH 12.5). Propanoic acid content ranged between 1.4 mg/l at pH 2 and 2.1 at pH 7; however, this acid was not detected in the extract at pH 12.5. Concentration of oxalic acid increased from 47.8 mg/l at pH 2, to 224.1 mg/l at pH 12.5, whereas the amount of malic acid varied between 8.6 and 14.1 mg/l at pH 2 and pH 12.5 respectively. MSA also increased from 1.1 (pH 2) to 6.1 mg/l (pH 12.5), but pyruvic acid was not observed. In the extracts of Waikato-1 coal with buffer solution at pH 2, hexanoic, heptanoic, octanoic and nonanoic acid were below the detection

[3]Although the pH in water extract varied after extraction, as convention the initial values will be used as reference on the text.

limit. Also, at pH 12.5 heptanoic, octanoic and nonanoic acid were below the detection limit. With the increase of pH, concentration of pentanoic acid varied as follows: 8.5, 61.3 and 12.6 µg/l at pH 2, 7 and 12.5 respectively. Hexanoic acid ranged between 42.2 µg/l (pH 7) and 8.9 µg/l (pH 12.5); whereas extracts at pH 7 exhibited concentrations of 33.4, 23.8 and 14.7 µg/l for heptanoic, octanoic and nonanoic acid respectively.

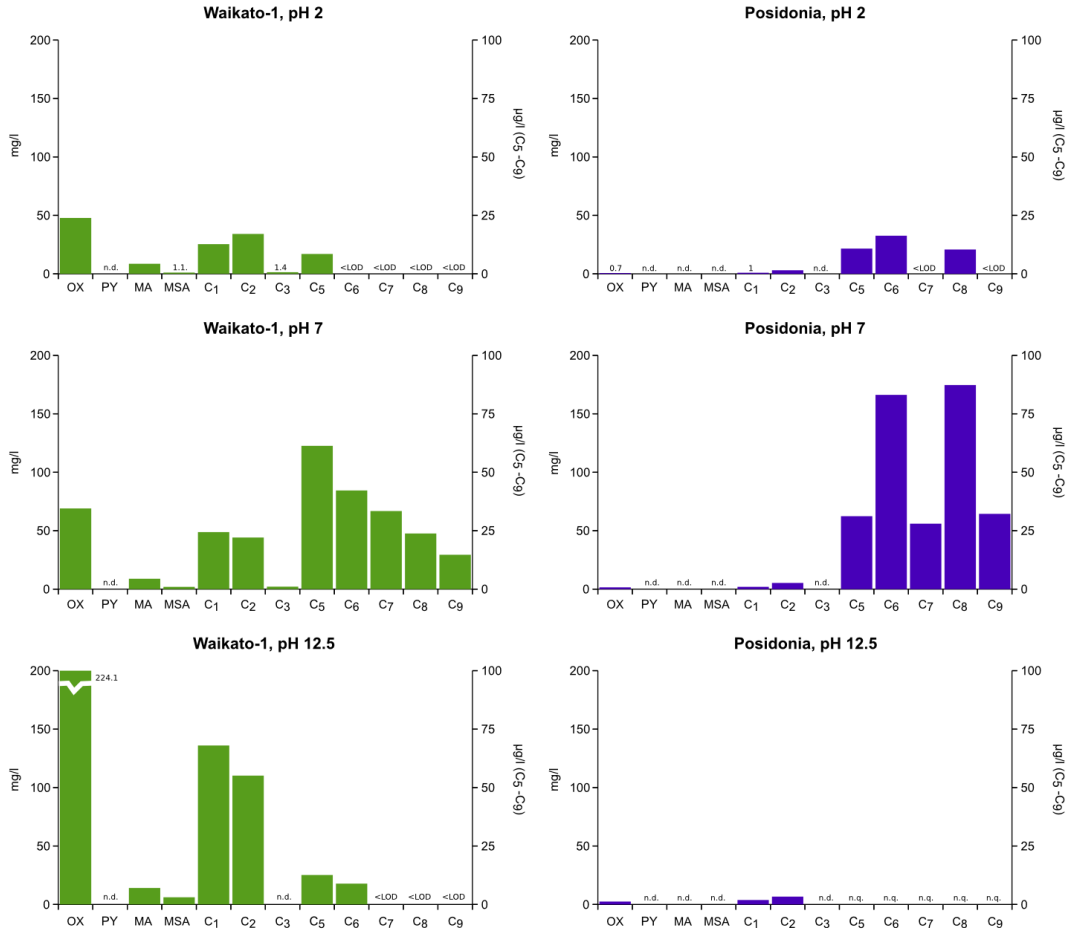


Fig. 6.16 Amounts of carboxylic acids in water extracts of Waikato-1 coal and Posidonia shale using buffered aqueous solutions at 2, 7 and 12.5 pH. For nomenclature, see Fig. 6.13; n.q. not quantified.

In the Posidonia shale extracts, with an increase of pH, concentrations of formic acid increased from 1 to 3.8 mg/l, while acetic acid content increased from 3.1 to 6.7 mg/l, and oxalic acid concentrations varied from 0.7 to 2.5 mg/l. Propionic, pyruvic and malic acid were not observed in the extracts. Pentanoic acid ranged as follows 10.8 µg/l at pH 2 and 31.2 µg/l at pH 7; whereas concentrations of hexanoic acid increased from 16.3 µg/l (pH 2) to 83.1 (pH 7) and content of octanoic acid varied from 10.4 µg/l (pH

2) to 87.3 µg/l (pH 7). Heptanoic and nonanoic acid were present above the limit of detection at pH 7, 28 and 32.2 µg/l respectively. Due to the possible precipitation of phosphates (described above), during the SPE preparation of the extract of Posidonia shale at pH 12.5, two dilutions were tested, 1:100 and 1:10. However, both dilutions were not suitable for quantifying acids with chain length from 5 up to 9 carbon number.

Summarizing, it could be seen that in the extracts of Waikato-1 coal the release of formic, acetic and oxalic acid was strongly enhanced when extraction was carried out under basic conditions; however, the extracts of Posidonia shale did not exhibit such a significant difference. Furthermore, in extracts of both samples concentrations of pentanoic, hexanoic, heptanoic, octanoic and nonanoic acid reached the highest value at pH 7.

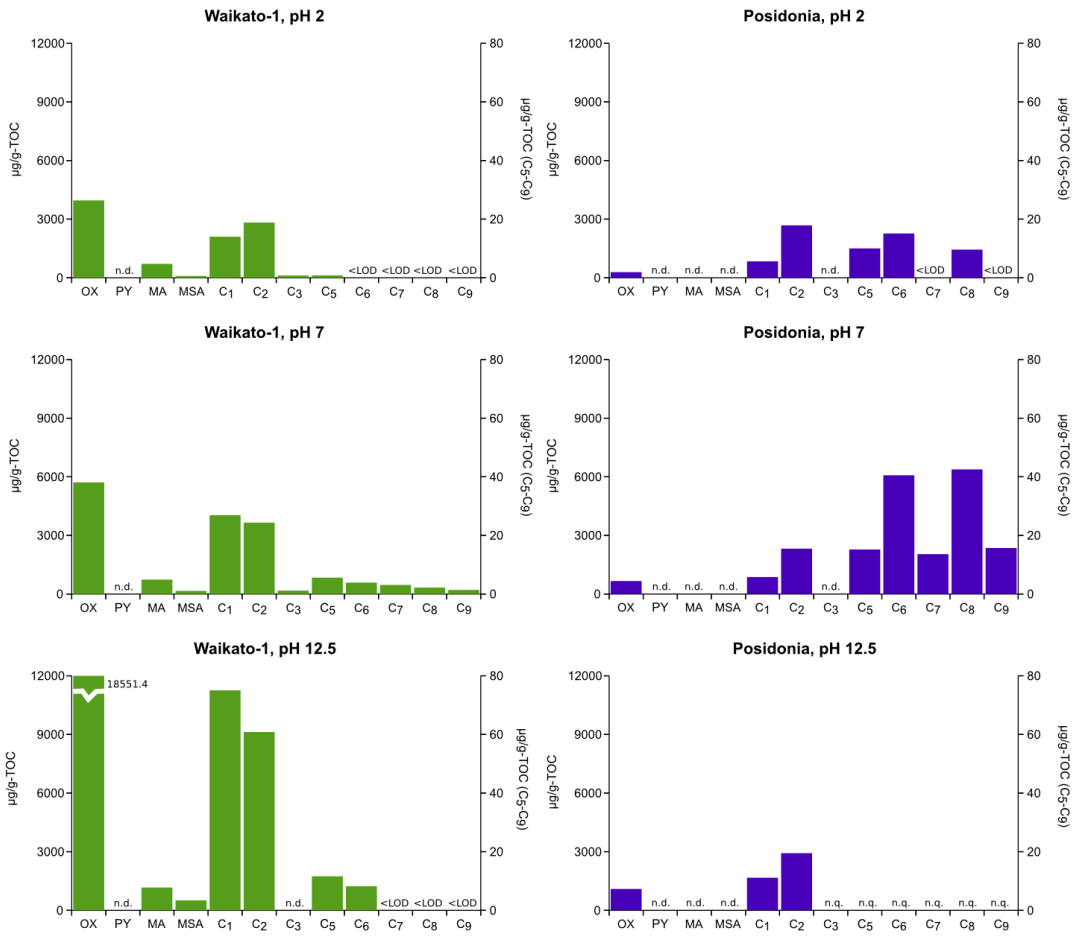


Fig. 6.17 Amounts of carboxylic acids, normalized to g-TOC, in the water extracts of Waikato-1 coal and Posidonia shale using buffered aqueous solutions at 2, 7 and 12.5 pH. For nomenclature see Fig. 6.13; n.q. not quantified.

Fig. 6.17 shows the amount of carboxylic acids normalized to TOC, i.e. µg/g-TOC, in the water extracts of Waikato-1 and Posidonia with acidic and basic phosphate-buffer solutions. From acidic to basic pH, in the extracts of Waikato-1, the content of formic acid increased from 2099.4 to 11258 µg/g-TOC, that of acetic acid varied from 2825 to 9122.5 µg/g-TOC, while the concentrations of propionic acid were from 114.4 (pH 2) to 174.5 µg/g-TOC (pH 7). In the extracts of Posidonia shale, concentrations of formic acid were as follows 840 µg/g-TOC at pH 2, 867 µg/g-TOC at pH 7 and 1666.7 at pH 12.5; whereas amounts of acetic acid were 2681.6 µg/g-TOC at pH 2, 2320.5 µg/g-TOC at pH 7, and 2924.4 µg/g-TOC at pH 12.5, and oxalic acid ranged from 286.9 (pH 2) to 1091.8 µg/g-TOC (pH 12.5). In the extracts of Waikato-1, an increase of pH, from 2 to 12.5, malic acid and MSA were observed at concentrations of 7711 up to 1165 µg/g-TOC and 89 up to 507 µg/g-TOC respectively.

In the aqueous extract of Waikato-1, an increase in pH raised the amounts of pentanoic acid from 0.8 to 11.6 µg/g-TOC; and concentrations from 3.9 to 15.1 µg/g-TOC for hexanoic acid. At pH 7 the concentration of heptanoic, octanoic and nonanoic acid were as follows 3.1, 2.2 and 1.4 µg/g-TOC respectively. In the extracts of Posidonia shale, an increase in pH changed the concentrations of pentanoic acid from 10 to 15.2 µg/g-TOC; while those of hexanoic varied from 15.1 to 40.5 µg/g-TOC, and yield of octanoic acid were between 9.6 to 42.5 µg/g-TOC. At pH 7, the amounts of heptanoic and nonanoic acid were 13.6 and 15.7 µg/g-TOC respectively. As already mentioned earlier, at pH 12.5 the concentrations of pentanoic to nonanoic acid were not quantified.

In the water extracts of the coal it could be seen that the concentrations of formic, acetic and oxalic increased significantly with an increase of pH, the amount of acids at pH 12.5 was at least 4 folds greater than those at pH 2. The aqueous extract of the shale did not exhibit such difference. For Waikato-1, it could also be seen that the amounts of formic and acetic acid extracted with a solution at pH 7 (water/sediment ratio of 50:1) are higher than those in the extracts with Millipore water (Fig. 6.14). Taking into account that pH are similar in both situations, but the only difference is the presence of phosphate; hence the variation may be due to mono- and dipotassium phosphate reduced the adsorption of the carboxylic acids onto the sediment, probably by increasing the particle dispersion or increasing the negative changes on both organic matter and sediment inorganic solid surfaces which amplify the repulsion amongst particles (Tani et al., 1993; You et al., 1999). Once again, the complex

solvation process, commented lines above, may makes these results possible.

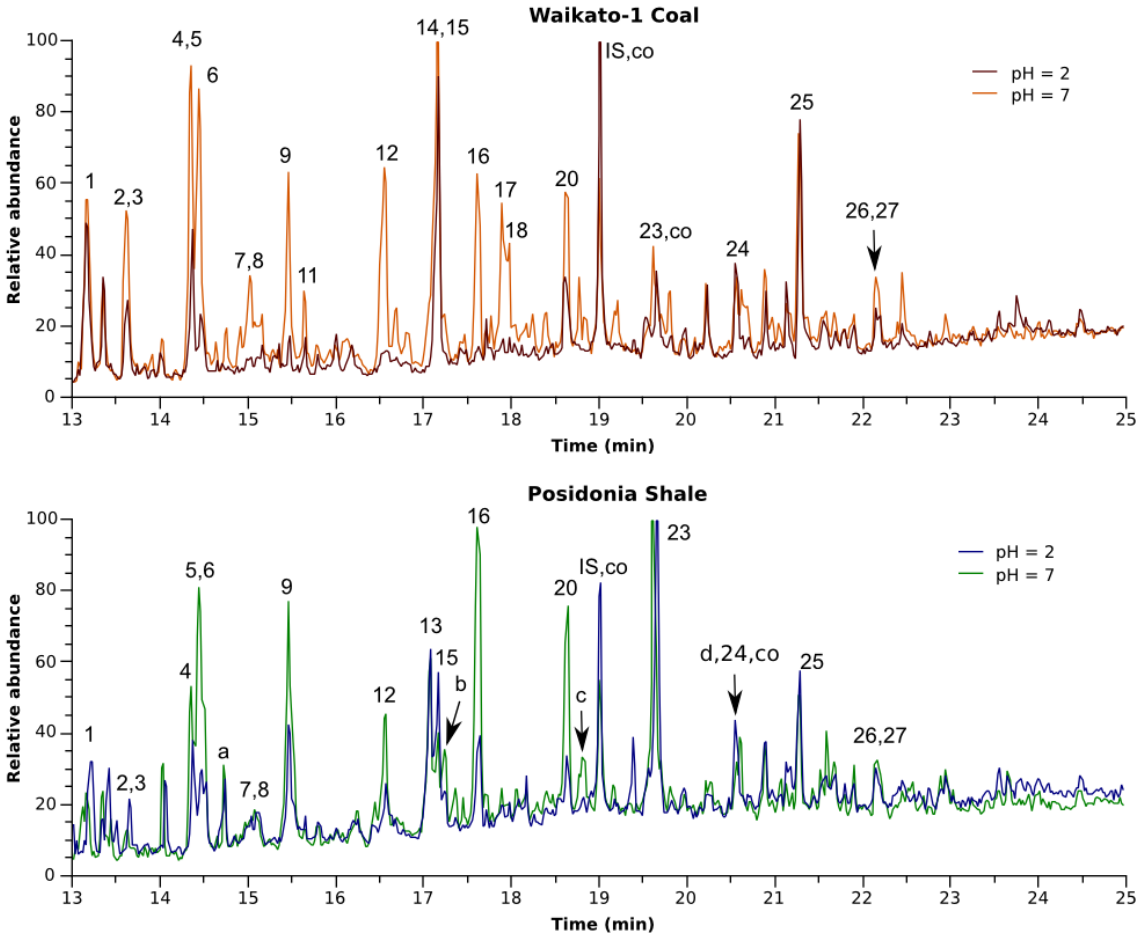


Fig. 6.18 GC-MS chromatograms of the SPE extracts of Waikato-1 coal and Posidonia shale under different pH conditions during aqueous extraction. For nomenclature see Table 6.1; a, 3-methylacetophenone; b, 1-indanone; c, 7-methyl-1-indanone; d, undecanoic acid.

6.7.3 Polar compounds dissolved after extraction with buffer

The diversity of polar compounds extracted with phosphate buffer was also examined (Fig 6.18). Comparing the substances released after extraction with Millipore water (Fig 6.1), GC-MS analysis showed that extracts (in acidified DMF) of Waikato-1 exhibited similar composition; whereas new compounds were observed in Posidonia shale extracts reported in Chapter 5.5. They have been tentatively identified as 3-methylacetophenone, 1-indanone, 7-methyl-1-indanone and undecanoic acid. Except for the latter, these compounds have been reported in low-polarity NSO compound fractions of Posidonia Shale bitumens by Wilkes *et al.*, (1998), as important structural types of aromatic carbonyl compounds; moreover, 1-indanones served to develop a heterocompound based maturity parameter. The authors proposed the geochemical

pathways required to form aromatic carbonyl compounds, namely inter- and/or intramolecular additions of acyl or formyl group to aromatic hydrocarbons. Alternatively the mechanism would be the oxidation of benzylic methylene groups of aromatic hydrocarbons.

Based on these results, extraction with phosphate buffer may enhance the release of compounds within the organic matter, likely by increasing the ionic strength i.e. reducing the attractive forces between ions. Organic-rich sediments may not naturally be exposed to high pH or phosphate buffer, but conditions where the media ionic strength is being altered are possible, consequently affecting the release of the acids. For example, in hazardous waste disposal facilities in geological formations where acidic leachates can be produced; or, during the process of hydraulic fracturing in unconventional reservoir stimulation which employs fluids containing acidic and oxidizing additives. The fluid-rock interactions under such conditions are still being investigated (Bildstein *et al.*, 2006; Moyce et al. 2014; Marcon *et al.*, 2017; Sutra *et al.*, 2017). Further, extraction of geological samples with this buffer may allow the examination of a wider range of water – soluble compounds using a mild temperature approach.

6.8 LMWOA as substrates for the deep biosphere

Calculation of generation rates is not feasible as the sample set considered in this study, does not represent a maturity series. However, a comparison between the concentrations of LMWOA determined in this study and those reported by Vieth *et al.,* (2008) and Glombitza *et al.*, (2009), can estimate the potential as substrates of the released LMWOA under aqueous extraction (the procedure employed in this study).

Table 6.11 Comparison of released LMWOA

Acid	Shales* mg/g TOC	Coals* mg/g TOC	Vieth *et al.*, (2008) mg/g TOC	Glombitza *et al.*, 2009 mg/g TOC
Formic acid	0-1.3	2.3-5.4	0.2-3.7	6.4-31.4
Acetic acid	0.2-1.9	1.2-4.4	0.2-1.4	1.5-6.5
Oxalic acid	0.1-1.4	1.1-3.5	0.2-1.4	0-5.7
*n*C5-*n*C9	0-0.3	0-0.01		

*Data (this study) only included from samples with R_0 < 0.55. *n*C5-*n*C9 represents the sum of concentrations of pentanoic to nonanoic acid.

Table 6.11 shows that the concentrations of formic, acetic and oxalic acid in 14 samples determined in this study, though lower, have similar order of magnitude to those reported in previous works. However, the amount of substrates that represent the range between pentanoic to nonanoic acid is much smaller compared to the concentrations of formic, acetic and oxalic acid. Tables 6.12 and 6.13 show the concentrations of LMWOA obtained by changing the extraction conditions (i.e. water/sediment ratio and pH-phosphate, see chapter 6.7) applied in this study, and those reported by Glombitza *et al.*, (2009). The comparison shows that an increase in the water/sediment ratio may provide very similar amounts of acetic and oxalic released from the whole sediment as the quantities that are liberated via ester-bound cleavage from isolated kerogen. Further, in the presence of potassium phosphate, the amounts of acetic and oxalic acid released are larger than those generated by ester-bound cleavage from isolated kerogen. Although this inorganic salt is not found naturally, this suggests that the amount of LMWOA contained in coals and shales is still large and could be released over time, thus representing an additional source of substrates for the deep biosphere. Although phosphate is common constituent of natural waters, it may also be present in the form of minerals (e.g. apatite). It is unclear what effect other forms of phosphate, and their interaction with cations (e.g. metals) can have on the release of LMWOA. Thus, a recommendation would be to carry out a new series of experiments testing the effect of some minerals (apatite) and cations on the amounts released of LMWOA.

Table 6.12 Comparison of released LMWOA – Water/sediment ratio

Acid	50:1 mg/g TOC	100:1 mg/g TOC	500:1 mg/g TOC	Glombitza et al. 2009 mg/g TOC
Formic acid	1.1-2.9	1.8-3.2	0-4.3	6.4-31.4
Acetic acid	2.3-2.4	2.5-2.6	2.7-3.5	1.5-6.5
Oxalic acid	0-1.2	0.3-2.8	0.3-6.4	0-5.7
*n*C5-*n*C9	0.008-0.15	0.01-0.15	0.02-0.17	

*n*C5-*n*C9 represents the sum of concentrations of pentanoic to nonanoic acid.

Table 6.13 Comparison of released LMWOA – pH - phosphate

Acid	pH 2 mg/g TOC	pH 7	PH 12.5	Glombitza et al. 2009
Formic acid	0.8-2.1	0.8-4	1.6-11.2	6.4-31.4
Acetic acid	2.6-2.8	2.3-3.6	2.9-9.1	1.5-6.5
Oxalic acid	0.2-3.9	0.6-5.7	1-18.5	0-5.7

*n*C5-*n*C9 represents the sum of concentrations of pentanoic to nonanoic acid.

Summary

Approximately 36 different substances as phenols, phenolic aldehydes and aliphatic, branched, aromatic, dicarboxylic, hydroxy and keto acids from one carbon chain length up to ten were identified in the water extracts of shales and coals. Amongst of them, 2-methoxyphenol (guaiacol) and 2,6-dimethoxyphenol (syringol) (lignin-derived components) and vanillin (biomarker of terrestrial matter), were found in some of the extracts of coals and in TS. Interestingly, 2-ethyl-3-methylmaleimide (EMMI), as biomarker for chlorophylls which may form under photic and/or oxygen rich sedimentary conditions, was detected in the extracts of all coals and in the extracts of Schöneck and Irati shales.

Pearson correlation showed that the acid release quantified by IC (µg/g-sed) was positively correlated with OI and TOC. A negative linear correlation was observed between acid content Tmax and Calc. Ro. When this coefficient was calculated using IC concentration normalized to TOC only OI exhibited a significant positive correlation, what suggests that OI may have a larger effect on the amount of released water-extractable acids than TOC and maturity. Multiple linear regression also showed that only OI may be used to predict the release of carboxylic acids from coals. Hierarchical cluster analysis applied on the concentrations quantified by IC (µg/g-sed), classified the samples in two groups: one conformed by Gore LM, Waikato, Karapiro, Waikato-1 and Lausitz; and other group comprised by the five shales and 2 coals (Cannel coal and Paparoa). The distinct feature amongst these subsets was the OI. Namely, the OI in the only-coal group ranged between 32 and 77 mgCO2/g-TOC; and in the most heterogeneous group, OI varied between 8 and 20 mgCO2/g-TOC.

Pearson correlation coefficient (using both µg/g-TOC and µg/g sed as units) between fluoride, chloride and sulfate and most abundant carboxylic acids was also calculated. Oxalic acid was more strongly related to fluoride and chloride.

The analysis of a wide range of LMWOA, from 1 to 9 carbon-chain number, allowed the identification of a pattern distribution in the acid content. In the extracts of coals, the amount of saturated acids decreased with an increase in the carbon chain, namely, the concentration of released acid decrease as follows: formic > acetic > pentanoic >

hexanoic > heptanoic > octanoic > nonanoic acid. This resembles a model where carboxylic acids are abiotically produced during oxidation of *n*-alkanes. However, this pattern was not observed in the extracts of shales. This suggested the

Two different extraction conditions were tested on Posidonia shale and Waikato-1 coal: water/sediment and pH-phosphate. The Waikato-1 coal released the highest amount of carboxylic acids per gram of TOC under the highest water/sediment ratio (500:1) in comparison to the extract of Posidonia shale. This may be explained by the combined effect of several processes (e.g. complex solvation). Further, it is suggested that it is related to the terrestrial matter of the coal as it may contain more interaction sites derived from oxygen-containing functional groups. Under this ratio (500:1), amongst the released acids, the increase observed in oxalic acid was remarkable. An increase of pH in the extraction solution from 2 to 7, resulted in an increase of the LMOWA released. Once again, the effect of pH in the release of oxalic acid was notable.

The amounts of LMWOA released under aqueous extraction have the same order of magnitude as those released by ester-bound cleavage to kerogen. Further, a larger amount of LMWOA can be released by simply increasing the water/sediment ratio, thus confirming their potential as feedstock lithologies for the deep biosphere. In the presence of potassium phosphate and alkaline pH, amounts of acetic and oxalic acid may exceed the concentrations of these acids liberated by ester-bound cleavage to kerogen. However, the effect of natural sources of phosphate (e.g. apatite) on the release of LMWOA should be evaluated.

7. Summary and Outlook

The potential application of LMWOA for societal benefit (e.g. improvements in exploitation of oil reserves, disposal of hazardous wastes and bioremediation), and as feasible substrate source for deep biosphere highlights the necessity of their detailed investigation. Many studies have quantified LMWOA with carbon chain length between 1 and 4. However, a detailed characterization of the water-soluble compounds that are released from geological samples is still missing. This study aimed to identify and quantify underivatized LMWOA that are released from coals and shales under subsurface conditions, i.e. in presence of water and 80-90 °C. In addition, the effects of water/sediment ratio, and pH – phosphate content on the release of LMWOA were evaluated.

The first task of this study was to design a procedure that isolated the acids from the aqueous matrix into an organic extract. Subsequently, this extract was planned to be analyzed by GC-MS to identify and to quantify the LMWOA. Several challenges identified during this procedure, and the steps taken to resolve them are summarized below.

The first methodological approach consisted of a series of **acid-base liquid-liquid extractions (LLE)** using dichloromethane (DCM) and diethyl ether (DEE) as solvents. The organic extract was then analyzed by GC-MS for identification and quantification of LMWOA. The first outcome of this test was that the target analytes (saturated carboxylic acids from propionic to dodecanoic acid) and the internal standard exhibited differences in sensitivity, in both GC-MS and GC-FID. The sensitivity was calculated for several branched and cyclic carboxylic acids with carbon atoms between 3 and 10, and they also exhibited this variability. It was clear that one-point calibration was inappropriate to quantify underivatized LMWOA using either GC-MS or GC-FID. Subsequently, relative response factors were calculated for each target analyte. The second, and more important, result of this test was the low recovery of the acid-base procedure. When the efficiency of the optimized applied acid-base LLE was assessed, only between 10 – 50 % of the tested carboxylic acids were recovered. Thus, acid-

base LLE was discarded for quantification of underivatized LMWOA by GC-MS.

Solid-phase extraction (SPE) was tested as the second methodological approach. In this method, GC-MS and GC-FID were applied to the organic extract for identification and quantification respectively. The setting of the SPE apparatus required some modifications until reproducible data (based on standard solutions) could be obtained. However, over time two problems arose. Firstly, unexpected peaks on the GC-FID chromatograms were observed when MeOH was used as eluent indicating that a reaction was occurring in the methanolic extract resulting in the degradation of the LMWOA. Several corrective measures were tested, substitution of the eluent being the most suitable solution. Thus, acidified DMF was chosen as new eluent. Once the optimized procedure was applied to natural samples, a new problem appeared. An unexpected hump was occasionally observed in the GC-FID chromatogram making the quantification unreliable. The cause of such a hump was found to be related to an excess of HCl (from the acidified sample), which reacted with the DMF. This time, the addition of a rinse step in the procedure resolved the problem.

The characterization of aqueous extracts involved the determination of DOC and its fractions, identification of water-soluble compounds and the quantification of underivatized carboxylic acids with carbon chain length from 1 up to 9. Correlation between DOC, concentrations of carboxylic acids and some inorganic anions was calculated via Pearson's coefficient. The results are summarized below:

1. In the **extracts of shales and coals**, weak negative correlations were found between DOC and Tmax, r = -0.75, R^2 = 0.56 and r = -0.71, R^2 = 0.51 respectively. In the extracts of shales, Pearson's r showed a weak negative correlation between DOC and vitrinite reflectance, r = -0.75, R^2 = 0.56; whereas DOC correlated strongly and negatively with the Suggate rank in coal samples. The lack of weakness of correlation has been interpreted as indicative of the absence of predominance of any of the assessed parameters over the release of the fraction that became dissolved into the aqueous matrix quantified as DOC. The extracts of coals were the exception, for those maturity (expressed in terms of the Suggate rank) showed a significant correlation with DOC. On a

HI vs Tmax diagram, no evident relationship was observed between the DOC in the extracts and kerogen type or geological age within the analyzed samples.

2. Amongst the series of oil-related samples, the extract of the **tailings sand** exhibited the highest DOC concentration and acid content. This result may be due to the TS is the result of several processes, from inorganic (e.g. dewatering by addition of polyacrylamide) to organic (e.g. plant growth) processes.
3. Approximately 36 compounds were identified in the aqueous extracts of shales and coals. This included phenols, phenolic aldehydes and aliphatic, branched, aromatic, dicarboxylic, hydroxy and keto acids. Additional compounds were found in the extracts. 2-methoxyphenol (guaiacol) and 2,6-dimethoxyphenol (syringol), lignin-derived components, were only identified in the extracts of coals. Interestingly, 2-ethyl-3-methylmaleimide (EMMI) was detected in extracts of the extracts of coals and Schöneck and Irati shales. This compound has been identified as biomarker for chlorophylls, which can be formed under photic and/or oxygenic sedimentary conditions.
4. With concentrations ranging between 182.4 and 1993.3 µg/g TOC, acetic acid was predominant in the extracts of **shales**, whereas, the formic acid content ranged between 35.2-1318 µg/g TOC. Generally, concentration of formic acid was higher (72.1-5470.2 µg/g TOC) in the extracts of coals, in comparison to the content of acetic acid. The *n*C5-*n*C9 fatty acid content was up to 322.6 µg/g-TOC, with the extracts of shales exhibiting higher yields in comparison to the extracts of coal (1.4-21.3 µg/g-TOC). When compared, the sum of formic and acetic acid is several orders of magnitude higher that the amount summing the content of pentanoic to nonanoic acid. In the extracts of shales, the oxalic acid content was between 107.8-1494.7 µg/g-TOC; whereas in the extracts of coals, the concentrations of this acid varied between 135.7-3539.7 µg/g-TOC.
5. The analysis of a wider range of LMWOA allowed the identification of a pattern distribution. In the extracts of coals, the amount of saturated acids decreased with an increase in the carbon chain, namely, the concentration of released acid decrease as follows: formic > acetic > pentanoic > hexanoic > heptanoic > octanoic > nonanoic acid. This resembles a model where carboxylic acids are abiotically produced during oxidation of *n*-alkanes. However, this pattern was not observed in the extracts of shales. Further, the extracts of Posedonia and

Schöneck shale showed a predominance of even-carbon carboxylic acid, which may be indicative of the biological activity in marine depositional environments. Briefly, the pattern distribution of saturated carboxylic acids seems to be dependent on both original organic matter and depositional environment.

6. Pearson coefficient showed that in the extracts of shales the correlation between acid content with OI was stronger than that with TOC. The correlation between the acid content and HI was moderately positive; whereas Pearson coefficient was moderately negative for TOC. Maturity indicators (Tmax and Calc. Ro) exhibited a weak correlation with the amount of acids. However, in extracts of coals, the acid content significantly correlated with OI. The Suggate Rank also showed strong correlation with the acid concentration, though it was negative. This observed linear correlation suggests that OI may have a larger effect on the amount of released water-extractable acid that TOC and maturity do.
7. A Student's t–test showed that the acid content averages in the extracts (expressed as µg/g-TOC) are significantly different for the two categories: shales and coals. This suggests that any equation to estimate the acid release should be defined separately, i.e. an equation for shales and another for coals.
8. Multiple linear regression defined a formula involving only OI as predictive variable, for estimation of the amounts of water-soluble LMWOA that can be released from immature coals.
9. Hierarchical cluster analysis applied to the acid content (quantified by IC, µg/g-sed) in the extracts of shales and coals, divided the samples in two groups: one conformed by Gore LM, Waikato, Karapiro, Waikato-1 and Lausitz; and, other group comprised by the five shales and 2 coals (Cannel coal and Paparoa). Interestingly, this division corresponded a classification based on the OI. Namely, the OI in the only-coal group ranged between 32 and 77 mgCO2/g-TOC; and in the heterogeneous group, OI varied between 8 and 20 mgCO2/g-TOC.
10. Pearson coefficient was also used to identify how formic, acetic and oxalic acid correlate with fluoride, chloride and sulfate. It was observed that oxalic acid was the compound more strongly related to inorganic ions, specifically fluoride and chloride. This may be due to oxalic acid is more reactive than formic and acetic

acid.

To evaluate circumstances that can impact the release of the acids, two conditions were tested: water/sediment ratio and pH-phosphate. This was applied to a coal and a shale samples. Three water/sediment ratios were tested by varying the amount of sediment. The pH-phosphate test was carried out using aqueous buffer solutions with pH 2, pH 7, and 12.5. The results showed that:

1. **Water/sediment ratio**: the extract of the coal yielded the highest amount of carboxylic acids per gram of TOC with the highest ratio (500:1), whereas the extracts of the shale did not exhibit strong variations. To some extent all acids exhibited an increase in yields with the highest water/sediment ratio (500:1), however the difference observed in oxalic acid was remarkable. This outcome could be explained by
    a. the more reactive nature of oxalic acid. Higher amount of water favored thedesorption of larger amount of oxalic acid. The adsorption of dicarboxylic acids as oxalic acid on sediments is stronger than the one of monocarboxylic acids as acetate;
    b. formation oxalate by oxidative cleavage at susceptible functional groups for instance 1,2-dihydroxycarboxylic, 1-hydroxy-2-keto- or 1,2-diketocarboxylic acid;
    c. the complex solvation mechanism that oxalic acid may undergo, as there are at least five known conformers for this diacid. Oxalic acid can form several cluster conformations depending on the number of water molecules, this influences the inter- and intramolecular hydrogen bonds formed, thus affecting the release of more oxalic acid molecules.

These results demonstrate the necessity for methodological standardization of quantification methods for comparison and use of data.

2. **pH-phosphate:** the increase in pH resulted in a significant increase in the concentrations of formic, acetic and oxalic acid in the extracts of the coal. In contrast, in the extracts of the shale, the increase in concentration due to an

increase of pH was very discrete. Oxalic acid being an exception. Once again, the content of oxalic acid underwent the most impressive change. Additional to the mechanisms that were mentioned lines above (explaining the effect of a high water/sediment ratio), the strong variation may be due to pH and mono- and dipotassium phosphate reduced the adsorption of the carboxylic acids onto the sediment, thus more amounts of this acid were dissolved during the aqueous extraction. These results shows the importance that dissolution of quelating minerals may have on release of LMWOA.

Summarizing, the present study shows that the released amounts of formic, acetic and oxalic acid are much larger than those of pentanoic, hexanoic, heptanoic, octanoic and nonanoic acid in the water extracts of shales and coals. Moreover, shales tend to release less amount of formic, acetic and oxalic acid in comparison to coals, but shales may provide significant amounts of pentanoic, hexanoic, heptanoic, octanoic and nonanoic depending on the depositional conditions. Sediments with high TOC, as coals, could be considered better feedstock for deep biosphere compared to shales; further, oxygen index may be used as only predictive variable to estimate the amount of LMWOA that are released. However, organic matter and maturity also play and important role in the release of LMWOA.

Future perspectives:

This study has not only resulted in the characterization of LMWOA released from geological samples but also demonstrated the challenges associated with the methodology in order to quantify underivatized carboxylic acids with chain containing between 5 and 9 carbon atoms. The technique used in this study has the potential to become a standard method that will allow faster and cheaper analysis of samples without use of hazardous chemicals. This will also allow inter-comparison of data from other investigations.

To complement the results of this study, a follow-up project should include:

- IC analysis of a wider range of coals and shales in order to develop an indicator based on OI to estimate the amount of LMWOA that can be released;

- testing of some minerals (e.g. apatite) and metals in the release of LMWOA;
- analysis of cations and dissolved inorganic carbon to identify how these parameters are related to the release of the LMWOA matter of the present study;
- inclusion of calibration curves to determine compounds as benzoic acid, 2,3-dimethylmaleic anhydride, 2-methoxyphenol (guaiacol), 2,6-dimethoxyphenol (syringol), 2-ethyl-3-methylmaleimide (EMMI). Some of these compounds have been used as indicators of vegetable matter or related to photosynthesis or another metabolic processes occurring under oxic or anoxic environments. Thus, their amounts could help to elucidate any synthesis/ degradation relationship between these compounds and LMWOA;
- organic and inorganic isotopic analysis in order to provide more information about the mechanisms that release the LMWOA from coals and shales, as questions about actual pathways and origin of LMWOA in coals and shales are still open.

8. References

Ábalos, M., Bayona, J. M., & Pawliszyn, J. 2000. Development of a headspace solid-phase microextraction procedure for the determination of free volatile fatty acids in waste waters. Journal of Chromatography A, 873(1), 107–115.

Ackman, R. G. 1964. Fundamental Groups in the Response of Flame Ionization Detectors to Oxygenated Aliphatic Hydrocarbons. Journal of Chromatographic Science, 2(6), 173–179.

Ackman, R. G., & Sipos, J. C. 1964. Flame ionization detector response for the carbonyl carbon atom in the carboxyl group of fatty acids and esters. Journal of Chromatography A, 16(0), 298–305.

Albarède, F., 2011. Oxygen Fugacity, in: Gargaud, M., Amils, R., Quintanilla, J.C., Cleaves, H.J. (Jim), Irvine, W.M., Pinti, D.L., Viso, M. (Eds.), Encyclopedia of Astrobiology. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 1196–1196. doi:10.1007/978-3-642-11274-4_4021

Barth, T., 1987. Quantitative determination of volatile carboxylic acids in formation waters by isotachophoresis. Analytical Chemistry 59, 2232-2237.

Barth, T., Borgund, A.E., Hopland, A.L., Graue, A., 1988. Volatile organic acids produced during kerogen maturation- amounts, composition and role in migration of oil. Organic Geochemistry 13, 461-465.

Becker, H. 1977. Organikum : organisch-chemisches Grundpraktikum ; mit 154 Tabellen. Berlin: Dt. Verl. d. Wiss.

Bernard, Sylvain, Brian Horsfield, Hans-Martin Schulz, Richard Wirth, Anja Schreiber, and Neil Sherwood. 2012. “Geochemical Evolution of Organic-Rich Shales with Increasing Maturity: A STXM and TEM Study of the Posidonia Shale (Lower Toarcian, Northern Germany).” Marine and Petroleum Geology 31 (1): 70–89.

Bildstein, O., Trotignon, L., Perronnet, M., Jullien, M., 2006. Modelling iron–clay interactions in deep geological disposal conditions. Migr. 2005 10th Int. Conf. Chem. Migr. Actin. Fission Prod. Geosph. 31, 618–625.

Blair, S.A., Thakkar, A.J., 2010. How many intramolecular hydrogen bonds does the oxalic acid dimer have? Chemical Physics Letters 495, 198-202.

Borgund, A.E., Barth, T., 1994. Generation of short-chain organic acids from crude oil

by hydrous pyrolysis. Organic Geochemistry 21, 943-952.

Bosak, T., Newman, D.K., 2005. Microbial Kinetic Controls on Calcite Morphology in Supersaturated Solutions. J. Sediment. Res. 75, 190–199.

Bou-Raad, M., Hobday, M. D., & Rix, C. J. (2000). Aqueous extraction of oxalate and other anions from coal. Fuel, 79(10), 1185–1193.

Bracewell, J.M., Robertson, G.W., Welch, D.I., 1980. Polycarboxylic acids as the origin of some pyrolysis products characteristics of soil organic matter. Journal of Analytical and Applied Pyrolysis 2, 239-248.

Bradley, J.-C., Neylon, C., Guha, R., Williams, A., Hooker, B., Lang, A., & Friesen, B. 2010. Open Notebook Science Challenge - Solubilities of Organic Compounds in Organic Solvents, 130.

Bradley, J.-C., Neylon, C., Guha, R., Williams, A., Hooker, B., Lang, A., & Friesen, B. 2015. Open Notebook Science Challenge - Solubilities of Organic Compounds in Organic Solvents. Retrieved Juli 10, 2015, from http://showme.physics.drexel.edu/onsc/models/SolubilityModel005.php?solutename=benzoic+acid&solutesmiles=c1ccccc1C%28=O%29O

Buchardt, B., Lewan, M.D., 1990. Reflectance of vitrinite-like macerals as a thermal maturity index for Cambrian-Ordovician Alum Shale, southern Scandinavia. AAPG Bull. 74, 394–406.

Carothers, W.W., Kharaka, Y.K., 1978. Aliphatic acid anions in oil-field waters; implications for origin of natural gas. AAPG Bulletin 62, 2441-2453.

Chebbi, A., Carlier, P., 1996. Carboxylic acids in the troposphere, occurrence, sources, and sinks: A review. Atmos. Environ. 30, 4233–4249.

Christie, W. W. 2011, July 8. Preparation of ester derivatives of fatty acids for chromatographic analysis. Retrieved January 1, 2012, from http://lipidlibrary.aocs.org/topics/ester_93/index.htm

Clark, J. 2012, September. Intermolecular bonding - van der Waals forces. Retrieved from http://www.chemguide.co.uk/atoms/bonding/vdw.html#top

Clark, J. 2014, August. The strengths of van der Waals dispersion forces. Retrieved from http://www.chemguide.co.uk/atoms/bonding/vdwstrengths.html#top

Correa Da Silva, Z.C., Cornford, C., 1985. The kerogen type, depositional environment and maturity, of the Irati Shale, Upper Permian of Paraná Basin,

Southern Brazil. Org. Geochem. 8, 399–411.

Cozzarelli, I.M., Baedecker, M.J., Eganhouse, R.P., Goerlitz, D.F., 1994. The geochemical evolution of low-molecular-weight organic acids derived from the degradation of petroleum contaminants in groundwater. Geochim. Cosmochim. Acta 58, 863–877.

Cuadros-Rodríguez, L., Bagur-González, M. G., Sánchez-Viñas, M., González-Casado, A., & Gómez-Sáez, A. M. 2007. Principles of analytical calibration/quantification for the separation sciences. Journal of Chromatography A, 1158(1–2), 33–46.

Cuadros-Rodríguez, L., Gámiz-Gracia, L., Almansa-López, E. M., & Bosque-Sendra, J. M. 2001. Calibration in chemical measurement processes. II. A methodological approach. TrAC Trends in Analytical Chemistry, 20(11), 620–636.

de Rijke, D., Traas, P. C., ter Heide, R., Boelens, H., & Takken, H. J. 1978. Acidic components in essential oils of costus root, patchouli and olibanum. Phytochemistry, 17(9), 1664–1666.

Dias, R. F., Freeman, K. H., Lewan, M. D., & Franks, S. G. 2002. [delta]13C of low-molecular-weight organic acids generated by the hydrous pyrolysis of oil-prone source rocks. Geochimica et Cosmochimica Acta, 66(15), 2755–2769.

Dhillon, A., Teske, A., Dillon, J., Stahl, D.A., Sogin, M.L., 2003. Molecular Characterization of Sulfate-Reducing Bacteria in the Guaymas Basin. Appl. Environ. Microbiol. 69, 2765–2772.

Duarte, B., Freitas, J., Caçador, I., 2011. The role of organic acids in assisted phytoremediation processes of salt marsh sediments. Hydrobiologia 1–9.

Edelkraut, F., 1996. Dissolved Vanillin as Tracer for Estuarine Lignin Conversion. Estuarine, Coastal and Shelf Science 43, 737-745.

Eglinton, T.I., Curtis, C.D., Rowland, S.J., 1987. Generation of Water-Soluble Organic Acids from Kerogen During Hydrous Pyrolysis: Implications for Porosity Development. Mineralogical Magazine 51, 495-503.

Gaylord, C. C. 2005. Reaction solvent Dimethyl sulfoxide (DMSO), 110.

Ghoos, Y., Geypens, B., Hiele, M., Rutgeerts, P., & Vantrappen, G. 1991. Analysis for short-chain carboxylic acids in feces by gas chromatography with an ion-trap

detector. Analytica Chimica Acta, 247(2), 223–227.

Glombitza, C., 2011. New Zealand coals - A potential feedstock for deep microbial life, Fakultät VI - Planen, Bauen, Umwelt. Technische Universität Berlin, Berlin, p. 337.

Glombitza, C., Mangelsdorf, K., Horsfield, B., 2009a. Maturation related changes in the distribution of ester bound fatty acids and alcohols in a coal series from the New Zealand Coal Band covering diagenetic to catagenetic coalification levels. Organic Geochemistry 40, 1063-1073.

Glombitza, C., Mangelsdorf, K., Horsfield, B., 2009b. A novel procedure to detect low molecular weight compounds released by alkaline ester cleavage from low maturity coals to assess its feedstock potential for deep microbial life. Organic Geochemistry 40, 175-183.

Gold, T., 1992. The deep, hot biosphere. Proc. Natl. Acad. Sci. 89, 6045–6049.

Goñi, M.A., Ruttenberg, K.C., Eglinton, T.I., 1997. Sources and contribution of terrigenous organic carbon to surface sediments in the Gulf of Mexico. Nature 389, 275-278.

Grice, K., Gibbison, R., Atkinson, J.E., Schwark, L., Eckardt, C.B., Maxwell, J.R., 1996. Maleimides (1H-pyrrole-2,5-diones) as molecular indicators of anoxygenic photosynthesis in ancient water columns. Geochimica et Cosmochimica Acta 60, 3913-3924.

Grigoryan, A.A., Cornish, S.L., Buziak, B., Lin, S., Cavallaro, A., Arensdorf, J.J., Voordouw, G., 2008. Competitive Oxidation of Volatile Fatty Acids by Sulfate- and Nitrate-Reducing Bacteria from an Oil Field in Argentina. Appl. Environ. Microbiol. 74, 4324–4335.

Harrison, A. G., Jones, E. G., Gupta, S. K., & Nagy, G. P. 1966. Total cross sections for ionization by electron impact. Canadian Journal of Chemistry, 44(16), 1967–1973.

Head, I.M., Jones, D.M., Larter, S.R., 2003. Biological activity in the deep subsurface and the origin of heavy oil. Nature 426, 344–352.

Helgeson, H.C., Knox, A.M., Owens, C.E., Shock, E.L., 1993. Petroleum, oil field waters, and authigenic mineral assemblages Are they in metastable equilibrium in hydrocarbon reservoirs. Geochim. Cosmochim. Acta 57, 3295–3339.

Hermida-Ramón, J.M., Cabaleiro-Lago, E.M., Rodríguez-Otero, J., 2004. Computational study of the dissociation of oxalic acid in water clusters. Chemical Physics 302, 53-60.

Hetényi, M., 1998. Oxygen index as an indicator of early organic maturity. Adv. Org. Geochem. 1997 Proc. 18th Int. Meet. Org. Geochem. Part Pet. Geochem. 29, 63–77.

Hibbert, D. B. 2007. Quality Assurance for the Analytical Chemistry Laboratory. Oxford University Press.

Holz, M., França, A.B., Souza, P.A., Iannuzzi, R., Rohn, R., 2010. A stratigraphic chart of the Late Carboniferous/Permian succession of the eastern border of the Paraná Basin, Brazil, South America. J. South Am. Earth Sci. 29, 381–399.

Huber, S.A., Balz, A., Abert, M., Pronk, W., 2011. Characterisation of aquatic humic and non-humic matter with size-exclusion chromatography – organic carbon detection – organic nitrogen detection (LC-OCD-OND). Water Research 45, 879-885.

Hubert, C., Voordouw, G., 2007. Oil Field Souring Control by Nitrate-Reducing Sulfurospirillum spp. That Outcompete Sulfate-Reducing Bacteria for Organic Electron Donors. Appl. Environ. Microbiol. 73, 2644–2652.

Ikan, R., Ioselis, P., Rubinsztain, Y., Aizenshtat, Z., Pugmire, R., Anderson, L.L., Ishiwatari, R., 1986. Carbohydrate origin of humic substances. Naturwissenschaften 73, 150-151.

Jensenius, J., 1987. Regional studies of fluid inclusions in Paleozoic sediments from southern Scandinavia. J. Name Bull Geol Soc Den Den. J. Vol. 36 Pages: 221–235.

Jensenius, J., 1987. Regional studies of fluid inclusions in Paleozoic sediments from southern Scandinavia. Journal Name: Bull. Geol. Soc. Den.; (Denmark); Journal Volume: 36, Pages: 221-235.

Jerkovic, I., Mastelic, J., 2001. Composition of Free and Glycosidically Bound Volatiles of Mentha aquatica L. Croatica Chemica Acta 74, 431-439.

Jones, D.L., 1998. Organic acids in the rhizosphere – a critical review. Plant and Soil 205, 25-44.

Jones, D.L., Dennis, P.G., Owen, A.G., van Hees, P.A.W., 2003. Organic acid

behavior in soils - misconceptions and knowledge gaps. Plant and Soil 248, 31-41.

Jurado-Sánchez, B., Ballesteros, E., & Gallego, M. 2010. Determination of carboxylic acids in water by gas chromatography using several detectors after flow preconcentration. Journal of Chromatography A, 1217(47), 7440–7447.

Kallmeyer, Jens, Kai Mangelsdorf, Barry Cragg, and Brian Horsfield. 2006. Techniques for Contamination Assessment During Drilling for Terrestrial Subsurface Sediments. Geomicrobiology Journal 23 (3–4): 227–39.

Kallmeyer, J., Pockalny, R., Adhikari, R. R., Smith, D. C., & D'Hondt, S. 2012. Global distribution of microbial abundance and biomass in subseafloor sediment. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(40), 16213–16216.

Kawamura, K., Ishiwatari, R., 1985. Conversion of sedimentary fatty acids from extractable (unbound + bound) to tightly bound form during mild heating. Org. Geochem. 8, 197–201.

Kawamura, K., Tannenbaum, E., Huizinga, B.J., Kaplan, I.R., 1986. Volatile organic acids generated from kerogen during laboratory heating. Geochemical Journal 20, 51-59.

Kawamura, K., Kaplan, I.R., 1987. Dicarboxylic acids generated by thermal alteration of kerogen and humic acids. Geochimica et Cosmochimica Acta 51, 3201-3207.

Kawamura, K., Usukura, K., 1993. Distributions of low molecular weight dicarboxylic acids in the North Pacific aerosol samples. J. Oceanogr. 49, 271–283.

Keene, W.C., Galloway, J.N., 1988. The biogeochemical cycling of formic and acetic acids through the troposphere: an overview of current understanding. Tellus B 40.

Kenkel, J. 2003. Analytical chemistry for technicians (Third edition). CRC Press.

Killops, S. D. and Killops, V. J. (2004). An introduction to organic geochemistry. Blackwell Publishing, Oxford, England.

Koskikallio, Jouko. 1956. "Kinetics of the Hydrolysis and Formation of Dimethylmaleic Anhydride in Solvent Mixtures." Acta Chemica Scandinavica 10: 822–30.

Kou, D., & Mitra, S. 2003. Extraction of Semivolatile Organic Compounds from Solid Matrices. In Sample Preparation Techniques in Analytical Chemistry (pp. 139–

182). John Wiley & Sons, Inc.

Kozono, M., Nomoto, S., Mita, H., Ishiwatari, R., Shimoyama, A., 2002. 2-Ethyl-3-methylmaleimide in Tokyo Bay Sediments Providing the First Evidence for its Formation from Chlorophylls in the Present Photic and Oxygenic Zone. Bioscience, Biotechnology, and Biochemistry 66, 1844-1847.

Kozono, M., Nomoto, S., Mita, H., Shimoyama, A., 2001. Detection of maleimides and their characteristics in Neogene sediments of the Shinjo basin, Japan. Geochem. J. 35, 225-236.

Krajewski, K.P., 2008. The Botneheia Formation (Middle Triassic) in Edgeøya and Barentsøya, Svalbard: lithostratigraphy, facies, phosphogenesis, paleoenvironment. Pol. Polar Res. 29, 319–364.

Krarup Pedersen, G., 1989. The sedimentology of lower palaeozoic black shales from the shallow wells Skelbro 1 and Billegrav 1, Bornholm, Denmark. Bull. Geol. Soc. Den. 37, 151–173.

Larsen, J. W., C. Islas-Flores, M. T. Aida, P. Opaprakasit, and P. Painter. 2005. “Kerogen Chemistry 2. Low-Temperature Anhydride Formation in Kerogens.” Energy Fuels 19: 145–51.

Littke, R., Leythaeuser, D., Rullkötter, J., Baker, D.R., 1991. Keys to the depositional history of the Posidonia Shale (Toarcian) in the Hils Syncline, northern Germany. Geol. Soc. Lond. Spec. Publ. 58, 311–333.

Lundegard, P.D., Senftle, J.T., 1987. Hydrous pyrolysis: a tool for the study of organic acid synthesis. Applied Geochemistry 2, 605-612.

MacGowan, D.B., Surdam, R.C., 1988. Difunctional carboxylic acid anions in oilfield waters. Organic Geochemistry 12, 245-259.

MacGowan, D.B., Surdam, R.C., 1990. Water-Rock Interactions Special Memorial Issue Ivan Barnes (1931–1989) Carboxylic acid anions in formation waters, San Joaquin Basin and Louisiana Gulf Coast, U.S.A.—Implications for clastic diagenesis. Applied Geochemistry 5, 687-701.

Machihara, T., Ishiwatari, R., 1983. Evaluation of alkaline permanganate oxidation method for the characterization of young kerogen. Organic Geochemistry 5, 111-119.

Mahlstedt, N. L. 2012. Evaluating the late gas potential of source rocks stemming

from different sedimentary environments. Technischen Universität Berlin, Berlin.

Mahlstedt, N., Horsfield, B., 2012. Metagenetic methane generation in gas shales I. Screening protocols using immature samples. Insights Shale Gas Explor. Exploit. 31, 27–42.

Marcon, V., Joseph, C., Carter, K.E., Hedges, S.W., Lopano, C.L., Guthrie, G.D., Hakala, J.A., 2017. Experimental insights into geochemical changes in hydraulically fractured Marcellus Shale. Appl. Geochem. 76, 36–50.

Markoš, J., Jelemenský, L., Šóoš, M., & Čamaj, V. (1999). Selective chlorination of acetone in the gas phase. Chemical Papers, 53(6), 8.

Mastelic, J., Milos, M., Kustrak, D., Radonic, A., 2000. Essential Oil and Glycosidically Bound Volatile Compounds from the Needles of Common Juniper (Juniperus communis L.). Croatica Chemica Acta 73, 585-593.

McDonald, J.H. 2014. Handbook of Biological Statistics (3rd ed.). Sparky House Publishing, Baltimore, Maryland.

McMaster, M. C. 2007a. Appendix B: GC/MS Troubleshooting Quick Reference. In GC/MS (pp. 159–164). John Wiley & Sons, Inc.

McMaster, M. C. 2007b. System Maintenance and Troubleshooting. In GC/MS (pp. 85–92). John Wiley & Sons, Inc.

McWilliam, I. G. 1983. The origin of the flame lonization detector. Chromatographia, 17(5), 241–243.

Merck-Millipore. 2015. LiChrolut® EN. Retrieved from http://www.emdmillipore.com/US/en/polymer-based-spe/_FCb.qB.ZGMAAAFCecZJ3_ww,nav

Milos, M., Radonic, A., 1996. Essential oil and glycosidically bound volatile compounds from Croatian Cupressus sempervirens L. Acta Pharmaceutica 46, 309-314.

Milos, M., Radonic, A., 2000. Gas chromatography mass spectral analysis of free and glycosidically bound volatile compounds from Juniperus oxycedrus L. growing wild in Croatia. Food Chemistry 68, 333-338.

Mitra, S. 2003. Sample Preparation Techniques in Analytical Chemistry. Hoboken: Wiley.

Mohler, F. L., Williamson, L., & Dean, H. M. (1950). Total Ionization of Hydrocarbons

From Mass Spectral Data. J. Res. Nat. Bur. Stds., 45(3), 235–238.

Morinaga, S., Ishiwatari, R., 1987. Gas chromatographic determination of C1–C5 low-molecular-weight organic acids in alkaline permanganate oxidation products of humic substances. Journal of Chromatography A 403, 225-231.

Mørk, A., Bjorøy, M., 1984. Mesozoic source rocks on Svalbard, in: Spencer, A.M. (Ed.), Petroleum Geology of the North European Margin. Springer Netherlands, pp. 371–382.

Mørk, A., Worsley, D., 2006. Triassic of Svalbard and the Barents Shelf, in: Boreal Triassic, NGF Abstracts and Proceedings of the Geological Society of Norway. Longyearbyen, Svalbard, p. 149 pp.

Moyce, E.B.A., Rochelle, C., Morris, K., Milodowski, A.E., Chen, X., Thornton, S., Small, J.S., Shaw, S., 2014. Rock alteration in alkaline cement waters over 15 years and its relevance to the geological disposal of nuclear waste. Appl. Geochem. 50, 91–105.

Naeher, S., Schaeffer, P., Adam, P., Schubert, C.J., 2013. Maleimides in recent sediments - Using chlorophyll degradation products for palaeoenvironmental reconstructions. Geochimica et Cosmochimica Acta 119, 248-263.

Nielsen, A.T., 1985. Carbon and oxygen isotope composition of Cambro-Silurian limestone and anthraconite from Bornholm: Evidence for deep burial diagenesis. Bulletin of the Geological Society of Denmark 33, 415-435.

Noah, M., Lappé, M., Schneider, B., Vieth-Hillebrand, A., Wilkes, H., Kallmeyer, J., 2014. Tracing biogeochemical and microbial variability over a complete oil sand mining and recultivation process. Science of The Total Environment 499, 297-310.

Noah, M., Poetz, S., Vieth-Hillebrand, A., & Wilkes, H. 2015. Detection of Residual Oil-Sand-Derived Organic Material in Developing Soils of Reclamation Sites by Ultra-High-Resolution Mass Spectrometry. Environmental Science & Technology, 49(11), 6466–6473.

Padula, V.T., 1969. Oil Shale of Permian Irati Formation, Brazil. AAPG Bull. 53, 591–602.

Pedersen, G.K., 1989. The sedimentology of Lower Palaeozoic black shales from the shallow wells Skelbro 1 and Billegrav 1, Bornholm, Denmark. Bull. Geol. Soc.

Den. 37, 151–173.

Pedersen, G.K., 1989. The sedimentology of Lower Palaeozoic black shales from the shallow wells Skelbro 1 and Billegrav 1, Bornholm, Denmark. Bulletin of the Geological Society of Denmark 37, 151-173.

Pedersen, K., 1997. Microbial life in deep granitic rock. FEMS Microbiol. Rev. 20, 399–414.

Pickering, W. F. 1985. The mobility of soluble fluoride in soils. Environmental Pollution Series B, Chemical and Physical, 9(4), 281–308.

Reichardt, C., & Welton, T. 2010. Appendix A. Properties, Purification, and Use of Organic Solvents. In Solvents and Solvent Effects in Organic Chemistry (pp. 549–586). Wiley-VCH Verlag GmbH & Co. KGaA.

Reichardt, C., Welton, T., 2011. Solvents and Solvent Effects in Organic Chemistry. Wiley-Blackwell.

Riis, F., Lundschien, B.A., Høy, T., Mørk, A., Mørk, M.B.E., 2008. Evolution of the Triassic shelf in the northern Barents Sea region. Pol. Polar Res. 27, 318–338.

Ringelberg, D. B., Sutton, S., & White, D. C. 1997. Biomass, bioactivity and biodiversity: microbial ecology of the deep subsurface: analysis of ester-linked phospholipid fatty acids. FEMS Microbiology Reviews, 20(3–4), 371–377.

Saiz-Jimenez, C., De Leeuw, J.W., 1984. Pyrolysis-gas chromatography-mass spectrometry of soil polysaccharides, soil fulvic acids and polymaleic acid. Organic Geochemistry 6, 287-293.

Saiz-Jimenez, C., De Leeuw, J.W., 1986. Lignin pyrolysis products: Their structures and their significance as biomarkers. Organic Geochemistry 10, 869-876.

Scanlon, J. T., & Willis, D. E. 1985. Calculation of Flame Ionization Detector Relative Response Factors Using the Effective Carbon Number Concept. Journal of Chromatographic Science, 23(8), 333–340.

Schovsbo, N.H., Nielsen, A.T., Klitten, K., Mathiesen, A., Rasmussen, P., 2011. Shale gas investigations in Denmark: Lower Palaeozoic shales on Bornholm. Geol. Surv. Den. Greenl. Bull. 23, 9–14.

Schulz, H.-M., Sachsenhofer, R.F., Bechtel, A., Polesny, H., Wagner, L., 2002. The origin of hydrocarbon source rocks in the Austrian Molasse Basin (Eocene–Oligocene transition). Mar. Pet. Geol. 19, 683–709.

Seifert, A., Brause, H., & Rascher, J. 1993. Geology of the Niederlausitz Lignite district, Germany. International Journal of Coal Geology, 23(1), 263–289.

Sigma-Aldrich Co. 1998. Bulletin 910. Guide to Solid Phase Extraction.

Sigma-Aldrich. n.d.. 2,4-hexadienoic acid. Retrieved from http://www.sigmaaldrich.com/catalog/product/aldrich/w392103

Stams, A.J.M., 1994. Metabolic interactions between anaerobic bacteria in methanogenic environments. Antonie Van Leeuwenhoek 66, 271–294.

Standke, G., Rascher, J., & Strauss, C. 1993. Relative sea-level fluctuations and brown coal formation around the Early-Middle Miocene boundary in the Lusatian Brown Coal District. Geologische Rundschau, 82(2), 295–305.

Stillinger, F.H., 1978. Proton Transfer Reactions and Kinetics in Water, in: Henry, E., Douglas, H. (Eds.), Theoretical Chemistry. Elsevier, pp. 177-234.

Strobel, B.W., 2001. Influence of vegetation on low-molecular-weight carboxylic acids in soil solution-a review. Geoderma 99, 169–198.

Stumm, W., 1992. Chemistry of the solid-water interface: Processes at the mineral-water and particle-water interface in natural systems. New York, NY (United States); John Wiley Sons, United States.

Sutra, E., Spada, M., Burgherr, P., 2017. Chemicals usage in stimulation processes for shale gas and deep geothermal systems: A comprehensive review and comparison. Renew. Sustain. Energy Rev. 77, 1–11.

Suzuki, N., Taguchi, K., 1984. Diagenesis of extractable and bound fatty acids in possible source rocks in Japan. Org. Geochem. 6, 125–133.

Tani, M., Higashi, T., Nagatsuka, S., 1993. Dynamics of Low-Molecular-Weight Aliphatic Carboxylic Acids (LACAs) in Forest Soils. Soil Sci. Plant Nutr. 39, 485-495.

Tani, M., Shida, K.S., Tsutsuki, K., Kondo, R., 2001. Determination of water-soluble low-molecular-weight organic acids in soils by ion chromatography. Soil Sci. Plant Nutr. 47, 387–397.

Tao, S., Lin, B., 2000. Water soluble organic carbon and its measurement in soil and sediment. Water Research 34, 1751-1755.

Thomas, J.D., 1997. The role of dissolved organic matter, particularly free amino acids and humic substances, in freshwater ecosystems. Freshwater Biology 38,

1-36.

Thurman, E., 1985. Aquatic Humic Substances, Organic geochemistry of natural waters. Nijhoff/Junk Publishers, Dodrecht, pp. 273-361.

Thurman, E., 1985a. Carboxylic Acids and Phenols, Organic geochemistry of natural waters. Nijhoff/Junk Publishers, Dodrecht, pp. 113-150.

Thurman, E., 1985b. Classification of dissolved organic carbon, Organic geochemistry of natural waters. Nijhoff/Junk Publishers, Dodrecht, pp. 103-112.

Vandenbroucke, M., Largeau, C., 2007. Kerogen origin, evolution and structure. Organic Geochemistry 38, 719-833.

Vieth, A., Mangelsdorf, K., Sykes, R., Horsfield, B., 2008. Water extraction of coals - potential for estimating low molecular weight organic acids as carbon feedstock for the deep terrestrial biosphere. Org. Geochem. 39, 985–991.

VIM. 2012. International Vocabulary of Metrology - Basic and General Concepts and Associated Terms (VIM) . 2012. Retrieved from http://www.bipm.org/utils/common/documents/jcgm/JCGM_200_2012.pdf

Vu, T.T.A., 2008. Origin and maturation of organic matter in New Zealand coals, Faculty of Mathematics and Natural Sciences. Ernst-Moritz-Arndt-University Greifswald, Greifswald, p. 268.

W. L. Lindsay, A.W.F.H.F.S., 1961. Identification of Reaction Products from Phosphate Fertilizers in Soils. Soil Science Society of America Journal 26, 446-452.

Watson, J. S., Jones, D. M., Swannell, R. P. J., & van Duin, A. C. T. 2002. Formation of carboxylic acids during aerobic biodegradation of crude oil and evidence of microbial oxidation of hopanes. Organic Geochemistry, 33(10), 1153–1169.

Wells, M. J. M. 2003. Principles of Extraction and the Extraction of Semivolatile Organics from Liquids. In Sample Preparation Techniques in Analytical Chemistry (pp. 37–138). John Wiley & Sons, Inc.

Wessa, P., (2016), Hierarchical Clustering (v1.0.4) in Free Statistics Software (v1.2.1), Office for Research Development and Education, URL http://www.wessa.net/rwasp_hierarchicalclustering.wasp/

Wilke, F. D. H., Vieth-Hillebrand, A., Naumann, R., Erzinger, J., & Horsfield, B. 2015. Induced mobility of inorganic and organic solutes from black shales using water

extraction: Implications for shale gas exploitation. Applied Geochemistry, 63, 158–168.

Wilkes, H., Kühner, S., Bolm, C., Fischer, T., Classen, A., Widdel, F., & Rabus, R. 2003. Formation of n-alkane- and cycloalkane-derived organic acids during anaerobic growth of a denitrifying bacterium with crude oil. Organic Geochemistry, 34(9), 1313–1323.

Yamabe, S., Tsuchida, N., 2003. A computational study of interactions between acetic acid and water molecules. Journal of Computational Chemistry 24, 939-947.

You, S.-J., Yin, Y., Allen, H.E., 1999. Partitioning of organic matter in soils: effects of pH and water/soil ratio. Science of The Total Environment 227, 155-160.

Appendix 1. Dissolved organic acids and its fractions by LC-OCD

Concentration (µg-C/g-sed)	Botneheia	Schöneck	Alum	Posidonia	Irati	Karapiro	Gore LM	Waikato-1	Waikato	Cannel Coal	Lausitz	Paparoa	OS	MFT	TS
DOC	229.7	580.9	27.6	432.8	265.9	6800.4	4503.7	1137.3	2754.5	424.8	4583.7	522.5	182.9	217.3	1544.2
HOC	120.2	171.6	n.d.	178.3	66.6	608.9	438.5	40.7	246.9	117	528.9	91.8	160.6	62.1	37.2
CDOC	109.4	409.3	27.6	254.6	199.3	6191.5	4065.2	1096.6	2507.6	307.8	4064.8	430.7	22.4	155.1	1504.47
HMW	0.6	1.2	3.5	0.8	1.4	136.4	51.1	3.4	6.2	5.8	41.2	11.9	0.7	9.1	96.8
CoRe	36.9	132.8	4.6	77.6	14.9	3963.1	2082.6	242.3	555	20.6	1204	53.9	4.6	41.3	1129.8
INT	28.6	96.1	1.9	50.7	25.7	1093.2	565.9	242.9	742.9	44.5	1108.9	133.5	5.9	53.6	133.6
LMW-neutras	35.6	123.6	16.2	65.2	101.8	911.3	657.3	118.4	174.5	55.2	714.6	141.3	9.5	50.3	145.7
LMW-acid	7.7	55.6	1.3	60.4	55.5	87.6	708.2	489.6	1028.9	181.8	996.1	467.1	1.5	0.7	n.d.

DOC, dissolved organic carbon; HOC, hydrophilic organic carbon; HMW, heaviest fraction (> 20000 Da); CoRe, chemically most complex and recalcitrant fraction (~ 1000 Da); INT, intermediate fraction which weight varies between 300 and 500 Da; LMW-acid and LMW-neutral fractions representing organic matter with MW lower than 350 Da; n.d., not detected.

Appendix 2. Concentration of carboxylic acids (µg/g-sed) – natural samples

Concentration (µg/g-sed)	Botneheia	Schöneck	Alum	Posidonia	Irati	Karapiro	Gore LM	Waikato-1	Waikato	Cannel Coal	Lausitz	Paparoa	OS	MFT	TS
Formic acid	28.6	< LOD	3.9	83.3	40.4	1232.8	2823.1	1807.2	1530.4	48.9	2569.5	247.6	< LOD	< LOD	< LOD
Acetic acid	4.9	188.5	20.4	227.2	186	553.1	2217.4	1428.3	2903.3	662.1	710	863.2	< LOD	n.d.	6
Propionic acid	n.d.	27.3	n.d.	14.8	27.9	4.3	95.2	164.9	225.9	45.9	42.3	< LOD	n.d.	n.d.	n.d.
Butanoic acid	n.d.	11.3	n.d.	12.1	6.5	n.d.	63.2	n.d.	51.5	n.d.	n.d.	< LOD	n.d.	n.d.	n.d.
Pentanoic acid	0.6	4.9	< LOD	2.4	1.8	1.3	4.9	2.6	5.8	0.6	4.2	1.6	0.4	1.2	0.7
Hexanoic acid	0.6	7.8	0.05	7.4	1.1	1.4	4.1	1	2.4	0.3	3	0.5	0.3	1	0.8
Heptanoic acid	0.5	6.8	< LOD	2	0.6	0.5	1.8	0.5	0.8	n.d.	1.1	< LOD	0.3	0.7	0.3
Octanoic acid	0.5	6.7	< LOD	6.5	0.4	0.4	0.3	0.6	0.2	n.d.	0.7	n.d.	0.3	0.7	0.7
Nonanoic acid	0.7	6.4	< LOD	2.2	n.d.	n.d.	n.d.	0.4	n.d.	n.d.	n.d.	n.d.	n.d.	0.6	n.d.
Oxalic acid	32.4	76	n.d.	66.4	27.3	1606.3	597.8	724.9	1776.4	92.1	1187.2	251.4	n.d.	58.2	142.9
Pyruvic acid	n.d.	n.d.	n.d.	n.d.	24.2	592.1	456.7	171	244	18.4	262.4	175.2	n.d.	n.d.	n.d.
Malic acid	n.d	n.d.	n.d.	n.d.	n.d.	146.5	419.9	204.7	245.2	10.6	131.3	< LOD	n.d.	n.d.	n.d.
Methylsuccinic acid	n.d	227.8	n.d.	n.d.	n.d.	33.9	144.8	82.4	117.9	n.d.	44	<LOD	n.d.	n.d.	n.d.

< LOD below limit of detection; n.d. no detected.

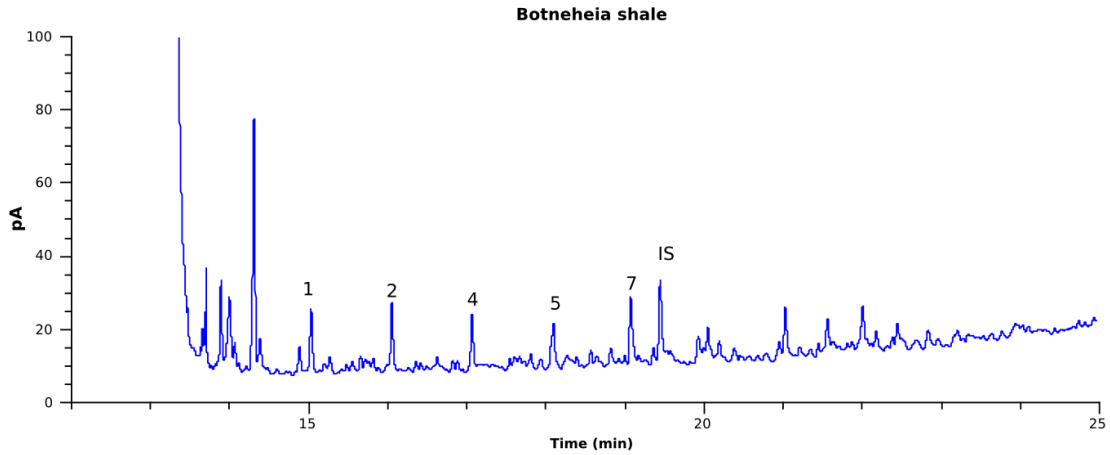
Appendix 3. Concentration of carboxylic acids - changing extraction conditions.

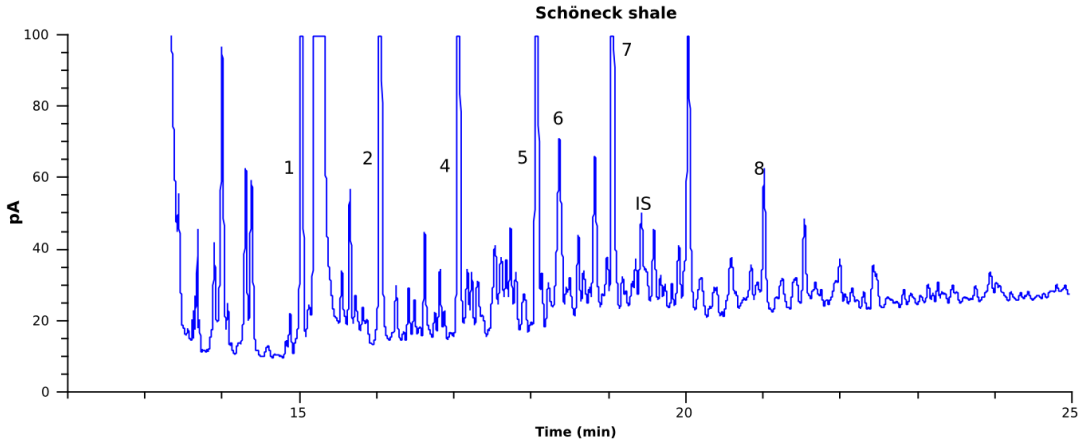
	Water/sediment ratio test									pH test					
	Waikato-1 coal						Posidonia shale			Waikato-1 coal			Posidonia shale		
Concentration (µg/g-sed)	**500:1**	**100:1**	**50:1**	**500:1***	**100:1***	**50:1***	**500:1**	**100:1**	**50:1**	**2**	**7**	**12.5**	**2**	**7**	**12.5**
Formic acid	2649.2	1980.4	1807.2	1682.9	1649.2	1478.3	273.1	136.2	125.2	1268.2	2438.5	6800.9	95.8	98.9	190
Acetic acid	2116.3	1616.9	1428.3	1969.5	1657.4	1498.8	316.6	286.5	280.5	1706.6	2207.5	5510.9	305.7	264.5	333.4
Propionic acid	174.5	179.1	164.9	150.8	166.9	174.2	n.d.	10.6	26.5	69.1	105.4	n.d.	n.d.	n.d.	n.d.
Pentanoic acid	5.1	3.4	2.6	n.q.			3.9	2.6	2.7	0.5	3.4	7	1.1	1.7	n.q.
Hexanoic acid	3.7	1.6	1				7.4	6.6	6.5	<LOD	2.3	4.9	1.7	4.6	n.q.
Heptanoic acid	3.4	1	0.5				<LOD	1.8	1.8	<LOD	1.9	<LOD	<LOD	1.6	n.q.
Octanoic acid	<LOD	1.2	0.6				7.6	6	6.2	<LOD	1.3	<LOD	1.1	4.9	n.q.
Nonanoic acid	<LOD	<LOD	0.4				4.5	2.8	2.6	<LOD	0.8	<LOD	<LOD	1.8	n.q.
Oxalic acid	3892.6	1739	724.9	2455.3	1365.9	720.2	n.d.	38	42.7	2390.7	3451.5	11206.9	32.7	75.9	124.5
Pyruvic acid	340.7	249.8	171	287.4	184.9	127.4	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
Malic acid	330.8	290.1	204.7	231.2	217.4	146.1	n.d.	n.d.	n.d.	429.8	445	704.1	n.d.	n.d.	n.d.
Methylsuccinic acid	<LOD	101.3	82.4	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	54.3	96.4	306.7	n.d.	n.d.	n.d.

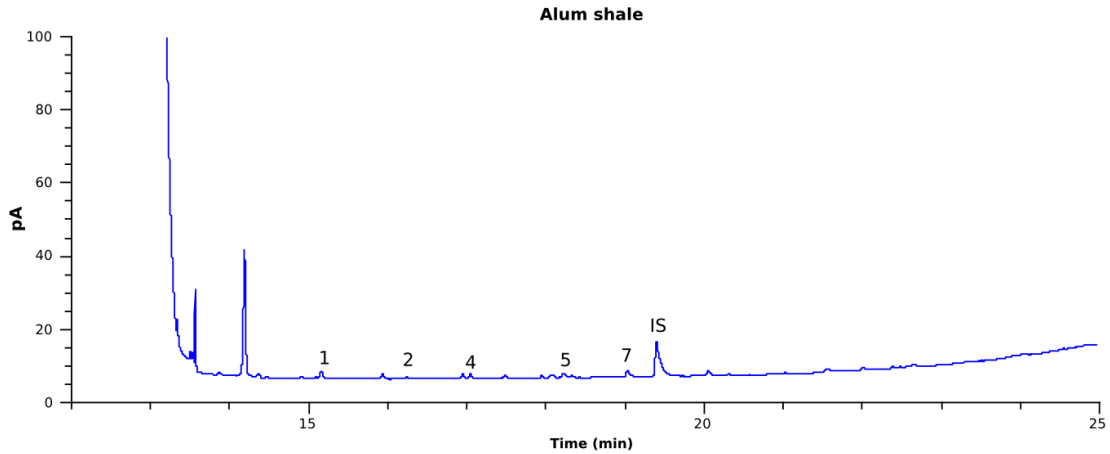
< LOD, below limit of detection; n.d. no detected; n.q., no quantified.

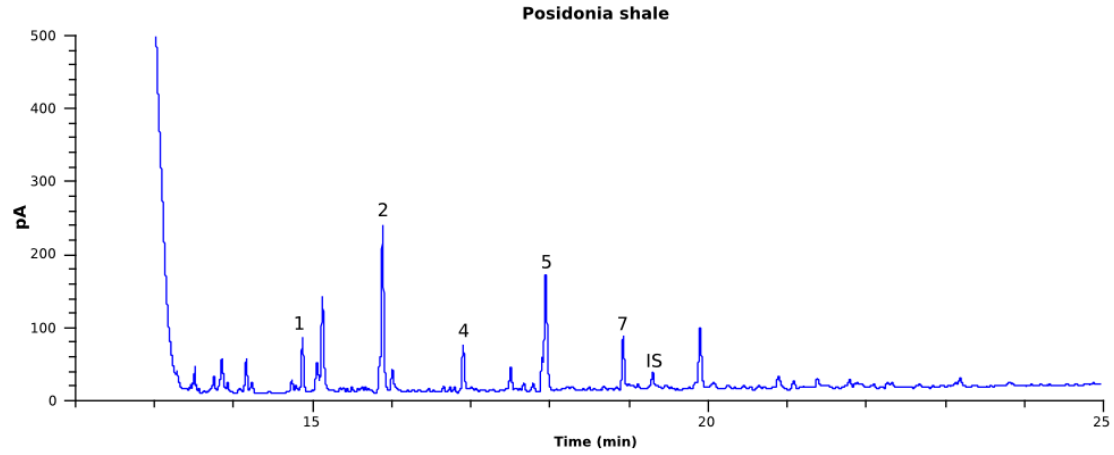
Appendix 4. GC-FID chromatograms of the SPE extracts of natural samples

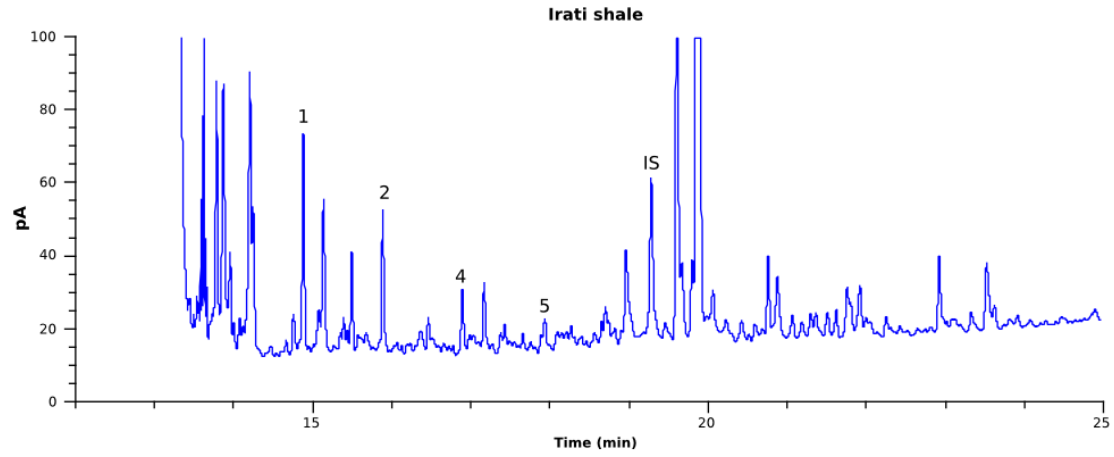
Nomenclature: 1, pentanoic acid; 2, hexanoic acid; 3, 2-methylhexanoic acid; 4, heptanoic acid; 5, octanoic acid; 6, cyclohexanoic acid; 7, nonanoic acid; 8, decanoic acid; IS, internal estandard – 4-ethyloctanoic acid.

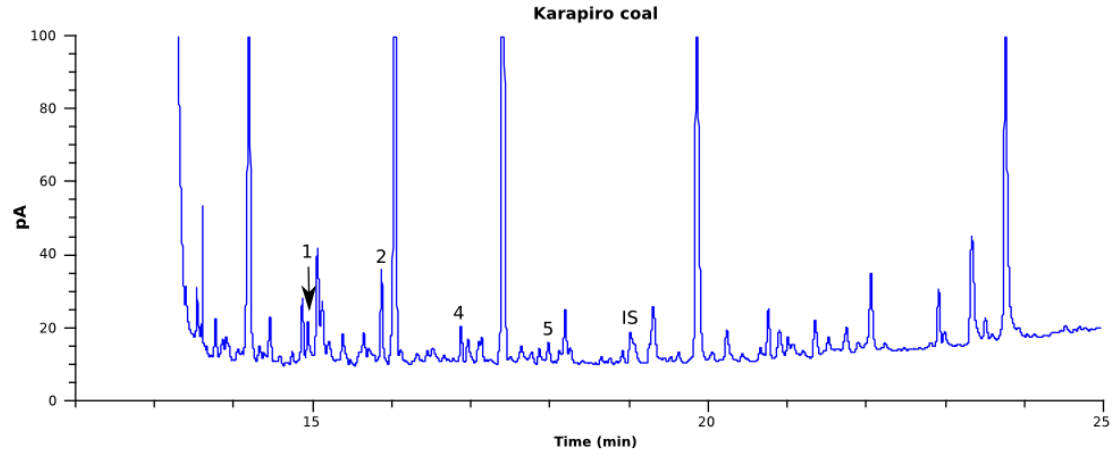


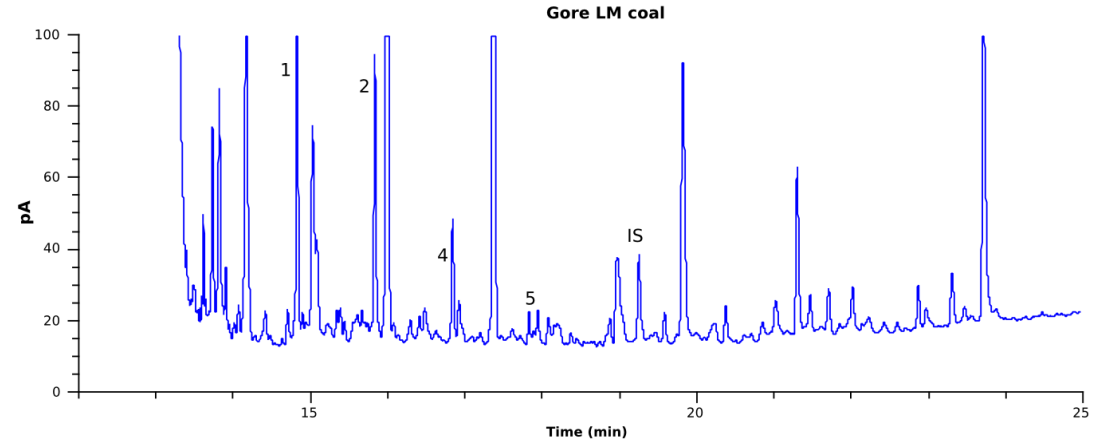


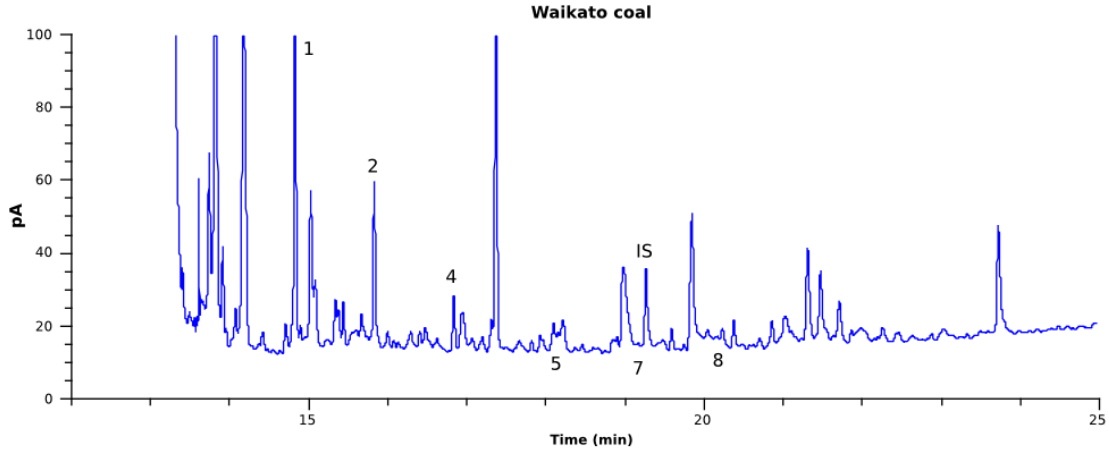


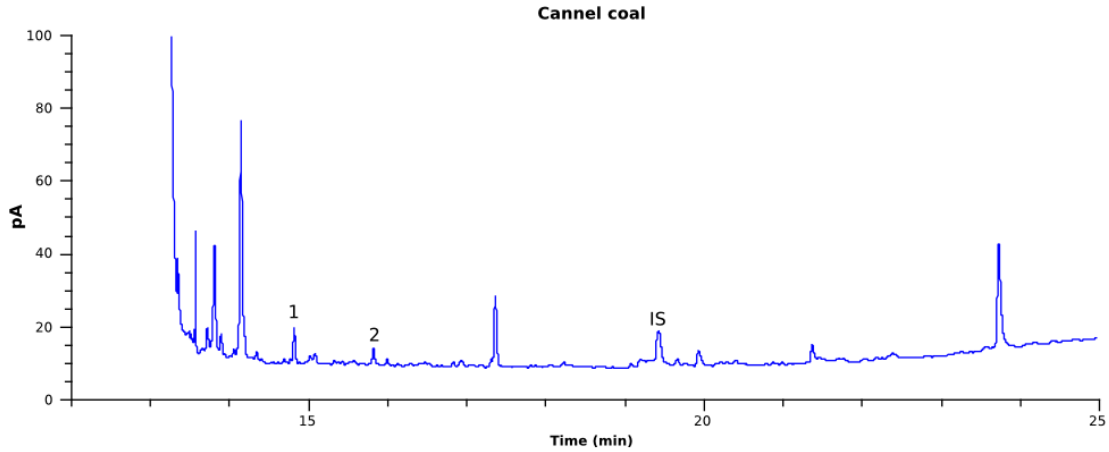

Posidonia shale
pA
Time (min)
1
2
4
5
7
IS

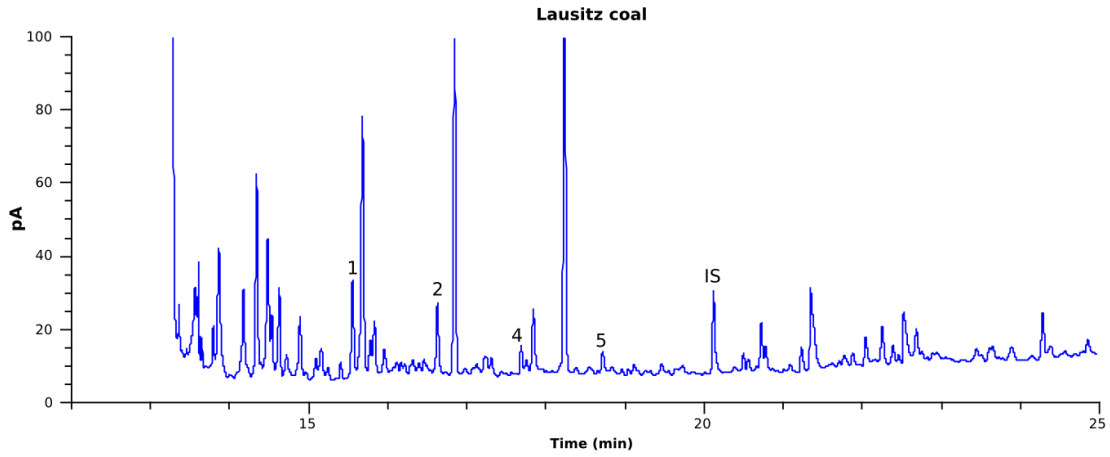

Irati shale
pA
Time (min)
1
2
4
5
IS

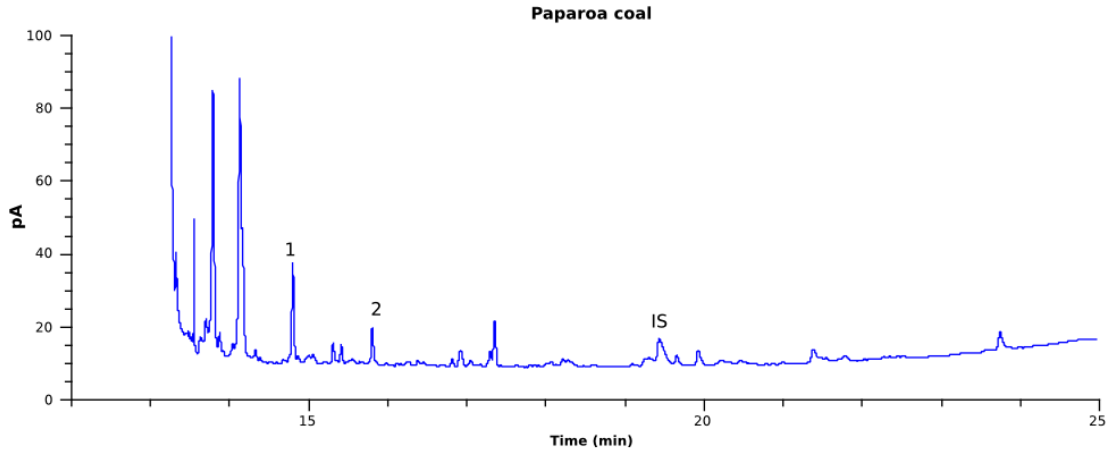

Karapiro coal
pA
Time (min)
1
2
4
5
IS

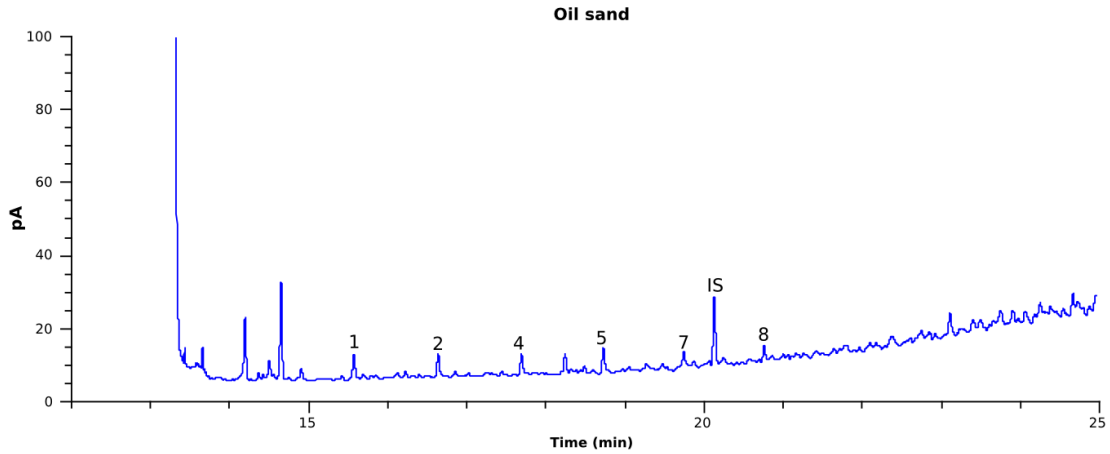

Gore LM coal
pA
Time (min)
1
2
4
5
IS

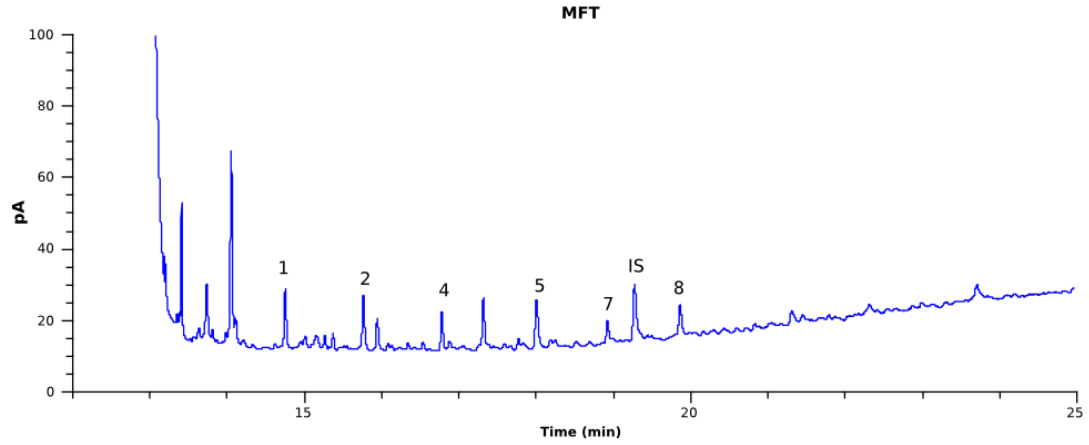

Waikato coal
pA
Time (min)
1
2
4
5
7
8
IS

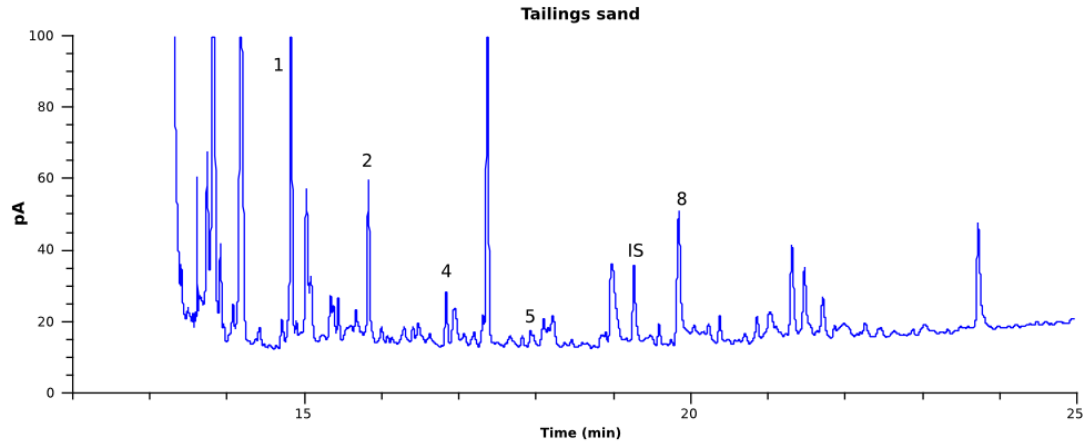

Cannel coal
pA
Time (min)
1
2
IS


Lausitz coal
pA
Time (min)
1
2
4
5
IS


Paparoa coal
pA
Time (min)
1
2
IS


Oil sand
pA
Time (min)
1
2
4
5
7
IS
8


MFT
pA
Time (min)
1
2
4
5
7
IS
8


Tailings sand
pA
Time (min)
1
2
4
5
IS
8

Appendix 5. Multiple regression

Multiple regression was performed using a spreadsheet by McDonald (2014). The following images show the procedure

1. Data entry

This workbook performs **multiple regression** with up to twelve independent (X) variables and up to 1000 observations.
For more information, see http://www.biostathandbook.com/multipleregression.html
It comes with the fish data from that page entered as an example.
To use the spreadsheet, start at this "data entry" sheet and replace the fish data with your numbers.
You may enter up to 1000 observations, with one dependent variable (Y-variable) and up to 12 independent variables (X-variables).
Enter the variable names in the appropriate row.
The observation labels in column A are optional.
If you want to use a bivariate nominal variable (such as "male" and "female") in your model, use the numerals 0 and 1.
After you've entered your data, go to the "univariate graphs" sheet (use the tabs at the bottom of this page) to see if you want to transform your data.

	Y-variable	X1	X2	X3	X4	X5	X6	X7	X8	X9
enter variable names--->	Acid conten	OI	TOC	Tmax						
Karapiro	4169	57	45.38	406						
Gore LM	6818.1	77	51.61	403						
Waikato-1	4583.4	37	60.41	407						
Waikato	7094.6	32	64.92	416						
Cannel Coal	878	8	67.9	430						
Lausitz	4946.7	65	69.4	392						
Paparoa	1537.4	6	74.18	431						

2. Preliminary equation

This sheet shows the results of the multiple regression.
Don't enter your data on this sheet; enter it on the "data entry" sheet.
Start with a "Y" above each variable name; this gives you a multiple regression using all the Y variables, as transformed on the "univariate graphs" sheet.
To do backward selection of X variables, look at the "P to remove" row. The X variable with the largest P-to-remove number has the least effect when remov
Put an "N" above the variable with the largest P-to-remove, then look at the new P-to-remove numbers.
Keep removing X variables until all of the P-to-remove values are below the cutoff you chose.
The coefficient of determination, P-value of the overall regression, and the regression equation are shown.

Once you've got a multiple regression you like, you can go to the "residuals graph" sheet to see how well it fits the data.

P to remove:	0.5923	0.9295	0.9943									
'N':	y	y	y	n	n	n	n	n	n	n	n	n
Y-variable name:	OI	TOC	Tmax									
coefficients:	67.2141	13.0506	1.4203									
number of X variables:	3											
determination (R^2):	0.53706											
P-value of R^2:	0.45285											
regression equation:	Acid content = +187.7042+67.21411*OI+13.0505*TOC+1.4202*Tmax											

3. Removing the variable with the highest P-value

P to remove:	0.1622	0.9096										
'N':	y	y	n	n	n	n	n	n	n	n	n	n
Y-variable name:	OI	TOC	Tmax									
coefficients:	66.4134	12.558										
number of X variables:	2											
determination (R^2):	0.53705											
P-value of R^2:	0.21432											
regression equation:	Acid content = +835.85150+66.41339*OI+12.5580*TOC											

4. Optimized equation using seven observations

P to remove:	0.0616											
'N':	y	n	n	n	n	n	n	n	n	n	n	n
Y-variable name:	OI	TOC	Tmax									
coefficients:	63.3582											
number of X variables:	1											
determination (R^2):	0.53536											
P-value of R^2:	0.0616											
regression equation:	Acid content = +1737.1713+63.35815*OI											

Appendix 6. Dendrograms

Dendrograms in text format with all details (R output)

Linkage rule = Single, 15 samples

```
R version 3.4.1 (2017-06-30) -- "Single Candle"

Copyright (C) 2017 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)


> source('/home/pw/wessanet/cretab')
>
>
>
> myrfcuid = ''
>
>                                                                 x                                                                 <-
array(list(65.9,530.9,24.3,403.8,312.3,4169,6818.1,4583.4,7094.6,878,4946.7,1537.4,0,58.2,148.9),di
m=c(1,15),dimnames=list(c('V1'),1:15))
>  y <- array(NA,dim=c(1,15),dimnames=list(c('V1'),1:15))
>  for (i in 1:dim(x)[1])
+  {
+    for (j in 1:dim(x)[2])
+    {
+         y[i,j] <- as.numeric(x[i,j])
+    }
+  }
> par4 = 'FALSE'
> par3 = 'FALSE'
> par2 = 'ALL'
> par1 = 'single'
> ylab = 'height'
> xlab = 'cases'
> main = 'Dendrogram'
> par4 <- 'FALSE'
> par3 <- 'FALSE'
> par2 <- 'ALL'
> par1 <- 'single'
> #'GNU S' R Code compiled by R2WASP v. 1.2.327 (Sun, 16 Jul 2017 10:31:41 +0200)
> #Author: root
> #To cite this work: Wessa, P., (2017), Hierarchical Clustering (v1.0.5) in Free Statistics Software
(v$_version), Office for Research Development and Education, URL
https://www.wessa.net/rwasp_hierarchicalclustering.wasp/
> #Source of accompanying publication: Office for Research, Development, and Education
> #
> par3 <- as.logical(par3)
> par4 <- as.logical(par4)
> if (par3 == 'TRUE'){
+ dum = xlab
+ xlab = ylab
+ ylab = dum
+ }
> x <- t(y)
> hc <- hclust(dist(x),method=par1)
> d <- as.dendrogram(hc)
> str(d)
--[dendrogram w/ 2 branches and 15 members at h = 2632]
  |--[dendrogram w/ 2 branches and 10 members at h = 659]
```

```
|  |--leaf "12"
|  `--[dendrogram w/ 2 branches and 9 members at h = 347]
|     |--leaf "10"
|     `--[dendrogram w/ 2 branches and 8 members at h = 163]
|        |--[dendrogram w/ 2 branches and 5 members at h = 83]
|        |  |--leaf "15"
|        |  `--[dendrogram w/ 2 branches and 4 members at h = 33.9]
|        |     |--[dendrogram w/ 2 branches and 2 members at h = 7.7]
|        |     |  |--leaf "1"
|        |     |  `--leaf "14"
|        |     `--[dendrogram w/ 2 branches and 2 members at h = 24.3]
|        |        |--leaf "3"
|        |        `--leaf "13"
|        `--[dendrogram w/ 2 branches and 3 members at h = 127]
|           |--leaf "2"
|           `--[dendrogram w/ 2 branches and 2 members at h = 91.5]
|              |--leaf "4"
|              `--leaf "5"
`--[dendrogram w/ 2 branches and 5 members at h = 1871]
   |--[dendrogram w/ 2 branches and 2 members at h = 276]
   |  |--leaf "7"
   |  `--leaf "9"
   `--[dendrogram w/ 2 branches and 3 members at h = 414]
      |--leaf "6"
      `--[dendrogram w/ 2 branches and 2 members at h = 363]
         |--leaf "8"
         `--leaf "11"
> mysub <- paste('Method: ',par1)
>
postscript(file="/home/pw/wessanet/rcomp/tmp/18v3s1500574991.ps",horizontal=F,onefile=F,pagecentre=F,paper="special",width=8.3333333333333,height=5.5555555555556)
> if (par4 == 'TRUE'){
+  plot(d,main=main,ylab=ylab,xlab=xlab,horiz=par3, nodePar=list(pch = c(1,NA), cex=0.8, lab.cex = 0.8),type='t',center=T, sub=mysub)
+ } else {
+  plot(d,main=main,ylab=ylab,xlab=xlab,horiz=par3, nodePar=list(pch = c(1,NA), cex=0.8, lab.cex = 0.8), sub=mysub)
+ }
> dev.off()
null device
          1
> if (par2 != 'ALL'){
+ if (par3 == 'TRUE'){
+ ylab = 'cluster'
+ } else {
+ xlab = 'cluster'
+ }
+ par2 <- as.numeric(par2)
+ memb <- cutree(hc, k = par2)
+ cent <- NULL
+ for(k in 1:par2){
+ cent <- rbind(cent, colMeans(x[memb == k, , drop = FALSE]))
+ }
+ hc1 <- hclust(dist(cent),method=par1, members = table(memb))
+ de <- as.dendrogram(hc1)
+
postscript(file="/home/pw/wessanet/rcomp/tmp/2flup1500574991.ps",horizontal=F,onefile=F,pagecentre=F,paper="special",width=8.3333333333333,height=5.5555555555556)
+ if (par4 == 'TRUE'){
+  plot(de,main=main,ylab=ylab,xlab=xlab,horiz=par3, nodePar=list(pch = c(1,NA), cex=0.8, lab.cex = 0.8),type='t',center=T, sub=mysub)
+ } else {
+  plot(de,main=main,ylab=ylab,xlab=xlab,horiz=par3, nodePar=list(pch = c(1,NA), cex=0.8, lab.cex =
```

```
0.8), sub=mysub)
+ }
+ dev.off()
+ str(de)
+ }
>
> a<-table.start()
> a<-table.row.start(a)
> a<-table.element(a,'Summary of Dendrogram',2,TRUE)
> a<-table.row.end(a)
> a<-table.row.start(a)
> a<-table.element(a,'Label',header=TRUE)
> a<-table.element(a,'Height',header=TRUE)
> a<-table.row.end(a)
> num <- length(x[,1])-1
> for (i in 1:num)
+ {
+ a<-table.row.start(a)
+ a<-table.element(a,hc$labels[i])
+ a<-table.element(a,hc$height[i])
+ a<-table.row.end(a)
+ }
> a<-table.end(a)
> table.save(a,file="/home/pw/wessanet/rcomp/tmp/3wado1500574991.tab")
> if (par2 != 'ALL'){
+ a<-table.start()
+ a<-table.row.start(a)
+ a<-table.element(a,'Summary of Cut Dendrogram',2,TRUE)
+ a<-table.row.end(a)
+ a<-table.row.start(a)
+ a<-table.element(a,'Label',header=TRUE)
+ a<-table.element(a,'Height',header=TRUE)
+ a<-table.row.end(a)
+ num <- par2-1
+ for (i in 1:num)
+ {
+ a<-table.row.start(a)
+ a<-table.element(a,i)
+ a<-table.element(a,hc1$height[i])
+ a<-table.row.end(a)
+ }
+ a<-table.end(a)
+ table.save(a,file="/home/pw/wessanet/rcomp/tmp/42a0a1500574991.tab")
+ }
>
> try(system("convert /home/pw/wessanet/rcomp/tmp/18v3s1500574991.ps
/home/pw/wessanet/rcomp/tmp/18v3s1500574991.png",intern=TRUE))
character(0)
> try(system("convert /home/pw/wessanet/rcomp/tmp/2flup1500574991.ps
/home/pw/wessanet/rcomp/tmp/2flup1500574991.png",intern=TRUE))
convert: unable to open image `/home/pw/wessanet/rcomp/tmp/2flup1500574991.ps': No such file or
directory @ error/blob.c/OpenBlob/2712.
convert: no images defined `/home/pw/wessanet/rcomp/tmp/2flup1500574991.png' @
error/convert.c/ConvertImageCommand/3210.
character(0)
attr(,"status")
[1] 1
Warning message:
running command 'convert /home/pw/wessanet/rcomp/tmp/2flup1500574991.ps
/home/pw/wessanet/rcomp/tmp/2flup1500574991.png' had status 1
>

> proc.time()
```

```
   user  system elapsed
  0.720   0.116   0.850
```

Linkage rule = Complete, 15 samples

```
R version 3.4.1 (2017-06-30) -- "Single Candle"

Copyright (C) 2017 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)

> source('/home/pw/wessanet/cretab')
>
>
>
> myrfcuid = ''
>
> x <-
array(list(65.9,530.9,24.3,403.8,312.3,4169,6818.1,4583.4,7094.6,878,4946.7,1537.4,0,58.2,148.9),di
m=c(1,15),dimnames=list(c('V1'),1:15))
>  y <- array(NA,dim=c(1,15),dimnames=list(c('V1'),1:15))
>  for (i in 1:dim(x)[1])
+  {
+       for (j in 1:dim(x)[2])
+       {
+               y[i,j] <- as.numeric(x[i,j])
+       }
+  }
> par4 = 'FALSE'
> par3 = 'FALSE'
> par2 = 'ALL'
> par1 = 'complete'
> ylab = 'height'
> xlab = 'cases'
> main = 'Dendrogram'
> par4 <- 'FALSE'
> par3 <- 'FALSE'
> par2 <- 'ALL'
> par1 <- 'complete'
> #'GNU S' R Code compiled by R2WASP v. 1.2.327 (Sun, 16 Jul 2017 10:31:41 +0200)
> #Author: root
> #To cite this work: Wessa, P., (2017), Hierarchical Clustering (v1.0.5) in Free Statistics Software
(v$_version), Office for Research Development and Education, URL
https://www.wessa.net/rwasp_hierarchicalclustering.wasp/
> #Source of accompanying publication: Office for Research, Development, and Education
> #
> par3 <- as.logical(par3)
> par4 <- as.logical(par4)
> if (par3 == 'TRUE'){
+ dum = xlab
+ xlab = ylab
+ ylab = dum
+ }
> x <- t(y)
> hc <- hclust(dist(x),method=par1)
> d <- as.dendrogram(hc)
> str(d)
--[dendrogram w/ 2 branches and 15 members at h = 7095]
  |--[dendrogram w/ 2 branches and 10 members at h = 1537]
  |  |--[dendrogram w/ 2 branches and 8 members at h = 531]
  |  |  |--[dendrogram w/ 2 branches and 5 members at h = 149]
```

```
  |  |  |  |--leaf "15"
  |  |  |  `--[dendrogram w/ 2 branches and 4 members at h = 65.9]
  |  |  |     |--[dendrogram w/ 2 branches and 2 members at h = 7.7]
  |  |  |     |  |--leaf "1"
  |  |  |     |  `--leaf "14"
  |  |  |     `--[dendrogram w/ 2 branches and 2 members at h = 24.3]
  |  |  |        |--leaf "3"
  |  |  |        `--leaf "13"
  |  |  `--[dendrogram w/ 2 branches and 3 members at h = 219]
  |  |     |--leaf "2"
  |  |     `--[dendrogram w/ 2 branches and 2 members at h = 91.5]
  |  |        |--leaf "4"
  |  |        `--leaf "5"
  |  `--[dendrogram w/ 2 branches and 2 members at h = 659]
  |     |--leaf "10"
  |     `--leaf "12"
  `--[dendrogram w/ 2 branches and 5 members at h = 2926]
     |--[dendrogram w/ 2 branches and 2 members at h = 276]
     |  |--leaf "7"
     |  `--leaf "9"
     `--[dendrogram w/ 2 branches and 3 members at h = 778]
        |--leaf "6"
        `--[dendrogram w/ 2 branches and 2 members at h = 363]
           |--leaf "8"
           `--leaf "11"
> mysub <- paste('Method: ',par1)
>
postscript(file="/home/pw/wessanet/rcomp/tmp/1y5xi1500506914.ps",horizontal=F,onefile=F,pagecentr
e=F,paper="special",width=8.3333333333333,height=5.5555555555556)
> if (par4 == 'TRUE'){
+ plot(d,main=main,ylab=ylab,xlab=xlab,horiz=par3, nodePar=list(pch = c(1,NA), cex=0.8, lab.cex =
0.8),type='t',center=T, sub=mysub)
+ } else {
+ plot(d,main=main,ylab=ylab,xlab=xlab,horiz=par3, nodePar=list(pch = c(1,NA), cex=0.8, lab.cex =
0.8), sub=mysub)
+ }
> dev.off()
null device
          1
> if (par2 != 'ALL'){
+ if (par3 == 'TRUE'){
+ ylab = 'cluster'
+ } else {
+ xlab = 'cluster'
+ }
+ par2 <- as.numeric(par2)
+ memb <- cutree(hc, k = par2)
+ cent <- NULL
+ for(k in 1:par2){
+ cent <- rbind(cent, colMeans(x[memb == k, , drop = FALSE]))
+ }
+ hc1 <- hclust(dist(cent),method=par1, members = table(memb))
+ de <- as.dendrogram(hc1)
+
postscript(file="/home/pw/wessanet/rcomp/tmp/2qavc1500506914.ps",horizontal=F,onefile=F,pagecent
re=F,paper="special",width=8.3333333333333,height=5.5555555555556)
+ if (par4 == 'TRUE'){
+ plot(de,main=main,ylab=ylab,xlab=xlab,horiz=par3, nodePar=list(pch = c(1,NA), cex=0.8, lab.cex =
0.8),type='t',center=T, sub=mysub)
+ } else {
+ plot(de,main=main,ylab=ylab,xlab=xlab,horiz=par3, nodePar=list(pch = c(1,NA), cex=0.8, lab.cex =
0.8), sub=mysub)
+ }
```

```
+ dev.off()
+ str(de)
+ }
>
> a<-table.start()
> a<-table.row.start(a)
> a<-table.element(a,'Summary of Dendrogram',2,TRUE)
> a<-table.row.end(a)
> a<-table.row.start(a)
> a<-table.element(a,'Label',header=TRUE)
> a<-table.element(a,'Height',header=TRUE)
> a<-table.row.end(a)
> num <- length(x[,1])-1
> for (i in 1:num)
+ {
+ a<-table.row.start(a)
+ a<-table.element(a,hc$labels[i])
+ a<-table.element(a,hc$height[i])
+ a<-table.row.end(a)
+ }
> a<-table.end(a)
> table.save(a,file="/home/pw/wessanet/rcomp/tmp/37amw1500506914.tab")
> if (par2 != 'ALL'){
+ a<-table.start()
+ a<-table.row.start(a)
+ a<-table.element(a,'Summary of Cut Dendrogram',2,TRUE)
+ a<-table.row.end(a)
+ a<-table.row.start(a)
+ a<-table.element(a,'Label',header=TRUE)
+ a<-table.element(a,'Height',header=TRUE)
+ a<-table.row.end(a)
+ num <- par2-1
+ for (i in 1:num)
+ {
+ a<-table.row.start(a)
+ a<-table.element(a,i)
+ a<-table.element(a,hc1$height[i])
+ a<-table.row.end(a)
+ }
+ a<-table.end(a)
+ table.save(a,file="/home/pw/wessanet/rcomp/tmp/46rgn1500506914.tab")
+ }
>
> try(system("convert /home/pw/wessanet/rcomp/tmp/1y5xi1500506914.ps
/home/pw/wessanet/rcomp/tmp/1y5xi1500506914.png",intern=TRUE))
character(0)
> try(system("convert /home/pw/wessanet/rcomp/tmp/2qavc1500506914.ps
/home/pw/wessanet/rcomp/tmp/2qavc1500506914.png",intern=TRUE))
convert: unable to open image `/home/pw/wessanet/rcomp/tmp/2qavc1500506914.ps': No such file or
directory @ error/blob.c/OpenBlob/2712.
convert: no images defined `/home/pw/wessanet/rcomp/tmp/2qavc1500506914.png' @
error/convert.c/ConvertImageCommand/3210.
character(0)
attr(,"status")
[1] 1
Warning message:
running command 'convert /home/pw/wessanet/rcomp/tmp/2qavc1500506914.ps
/home/pw/wessanet/rcomp/tmp/2qavc1500506914.png' had status 1
>
> proc.time()
   user  system elapsed
  0.616   0.096   0.720
```

Raw output Dendogram – Linkage rule = Complete, 12 samples

```
R version 3.4.1 (2017-06-30) -- "Single Candle"

Copyright (C) 2017 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)

> source('/home/pw/wessanet/cretab')
>
>
>
> myrfcuid = ''
>
> x <-
array(list(65.9,530.9,24.3,403.8,312.3,4169,6818.1,4583.4,7094.6,878,4946.7,1537.4),dim=c(1,12),di
mnames=list(c('V1'),1:12))
>  y <- array(NA,dim=c(1,12),dimnames=list(c('V1'),1:12))
>  for (i in 1:dim(x)[1])
+  {
+    for (j in 1:dim(x)[2])
+    {
+         y[i,j] <- as.numeric(x[i,j])
+    }
+  }
> par4 = 'FALSE'
> par3 = 'FALSE'
> par2 = 'ALL'
> par1 = 'complete'
> ylab = 'height'
> xlab = 'cases'
> main = 'Dendrogram'
> par4 <- 'FALSE'
> par3 <- 'FALSE'
> par2 <- 'ALL'
> par1 <- 'complete'
> #'GNU S' R Code compiled by R2WASP v. 1.2.327 (Sun, 16 Jul 2017 10:31:41 +0200)
> #Author: root
> #To cite this work: Wessa, P., (2017), Hierarchical Clustering (v1.0.5) in Free Statistics Software
(v$_version), Office for Research Development and Education, URL
https://www.wessa.net/rwasp_hierarchicalclustering.wasp/
> #Source of accompanying publication: Office for Research, Development, and Education
> #
> par3 <- as.logical(par3)
> par4 <- as.logical(par4)
> if (par3 == 'TRUE'){
+ dum = xlab
+ xlab = ylab
+ ylab = dum
+ }
> x <- t(y)
> hc <- hclust(dist(x),method=par1)
> d <- as.dendrogram(hc)
> str(d)
--[dendrogram w/ 2 branches and 12 members at h = 7070]
  |--[dendrogram w/ 2 branches and 7 members at h = 1513]
  |  |--[dendrogram w/ 2 branches and 5 members at h = 507]
  |  |  |--[dendrogram w/ 2 branches and 2 members at h = 41.6]
  |  |  |  |--leaf "1"
  |  |  |  `--leaf "3"
  |  |  `--[dendrogram w/ 2 branches and 3 members at h = 219]
  |  |     |--leaf "2"
  |  |     `--[dendrogram w/ 2 branches and 2 members at h = 91.5]
```

```
 |  |       |--leaf "4"
 |  |       `--leaf "5"
 |  `--[dendrogram w/ 2 branches and 2 members at h = 659]
 |     |--leaf "10"
 |     `--leaf "12"
 `--[dendrogram w/ 2 branches and 5 members at h = 2926]
    |--[dendrogram w/ 2 branches and 2 members at h = 276]
    |  |--leaf "7"
    |  `--leaf "9"
    `--[dendrogram w/ 2 branches and 3 members at h = 778]
       |--leaf "6"
       `--[dendrogram w/ 2 branches and 2 members at h = 363]
          |--leaf "8"
          `--leaf "11"
> mysub <- paste('Method: ',par1)
>
postscript(file="/home/pw/wessanet/rcomp/tmp/1k0ju1500576137.ps",horizontal=F,onefile=F,pagecentr
e=F,paper="special",width=8.3333333333333,height=5.5555555555556)
> if (par4 == 'TRUE'){
+ plot(d,main=main,ylab=ylab,xlab=xlab,horiz=par3, nodePar=list(pch = c(1,NA), cex=0.8, lab.cex =
0.8),type='t',center=T, sub=mysub)
+ } else {
+ plot(d,main=main,ylab=ylab,xlab=xlab,horiz=par3, nodePar=list(pch = c(1,NA), cex=0.8, lab.cex =
0.8), sub=mysub)
+ }
> dev.off()
null device
          1
> if (par2 != 'ALL'){
+ if (par3 == 'TRUE'){
+ ylab = 'cluster'
+ } else {
+ xlab = 'cluster'
+ }
+ par2 <- as.numeric(par2)
+ memb <- cutree(hc, k = par2)
+ cent <- NULL
+ for(k in 1:par2){
+ cent <- rbind(cent, colMeans(x[memb == k, , drop = FALSE]))
+ }
+ hc1 <- hclust(dist(cent),method=par1, members = table(memb))
+ de <- as.dendrogram(hc1)
+
postscript(file="/home/pw/wessanet/rcomp/tmp/2ny3n1500576137.ps",horizontal=F,onefile=F,pagecent
re=F,paper="special",width=8.3333333333333,height=5.5555555555556)
+ if (par4 == 'TRUE'){
+ plot(de,main=main,ylab=ylab,xlab=xlab,horiz=par3, nodePar=list(pch = c(1,NA), cex=0.8, lab.cex =
0.8),type='t',center=T, sub=mysub)
+ } else {
+ plot(de,main=main,ylab=ylab,xlab=xlab,horiz=par3, nodePar=list(pch = c(1,NA), cex=0.8, lab.cex =
0.8), sub=mysub)
+ }
+ dev.off()
+ str(de)
+ }
>
> a<-table.start()
> a<-table.row.start(a)
> a<-table.element(a,'Summary of Dendrogram',2,TRUE)
> a<-table.row.end(a)
> a<-table.row.start(a)
> a<-table.element(a,'Label',header=TRUE)
> a<-table.element(a,'Height',header=TRUE)
```

```
> a<-table.row.end(a)
> num <- length(x[,1])-1
> for (i in 1:num)
+ {
+ a<-table.row.start(a)
+ a<-table.element(a,hc$labels[i])
+ a<-table.element(a,hc$height[i])
+ a<-table.row.end(a)
+ }
> a<-table.end(a)
> table.save(a,file="/home/pw/wessanet/rcomp/tmp/35rls1500576137.tab")
> if (par2 != 'ALL'){
+ a<-table.start()
+ a<-table.row.start(a)
+ a<-table.element(a,'Summary of Cut Dendrogram',2,TRUE)
+ a<-table.row.end(a)
+ a<-table.row.start(a)
+ a<-table.element(a,'Label',header=TRUE)
+ a<-table.element(a,'Height',header=TRUE)
+ a<-table.row.end(a)
+ num <- par2-1
+ for (i in 1:num)
+ {
+ a<-table.row.start(a)
+ a<-table.element(a,i)
+ a<-table.element(a,hc1$height[i])
+ a<-table.row.end(a)
+ }
+ a<-table.end(a)
+ table.save(a,file="/home/pw/wessanet/rcomp/tmp/4uu541500576137.tab")
+ }
>
> try(system("convert /home/pw/wessanet/rcomp/tmp/1k0ju1500576137.ps
/home/pw/wessanet/rcomp/tmp/1k0ju1500576137.png",intern=TRUE))
character(0)
> try(system("convert /home/pw/wessanet/rcomp/tmp/2ny3n1500576137.ps
/home/pw/wessanet/rcomp/tmp/2ny3n1500576137.png",intern=TRUE))
convert: unable to open image `/home/pw/wessanet/rcomp/tmp/2ny3n1500576137.ps': No such file or
directory @ error/blob.c/OpenBlob/2712.
convert: no images defined `/home/pw/wessanet/rcomp/tmp/2ny3n1500576137.png' @
error/convert.c/ConvertImageCommand/3210.
character(0)
attr(,"status")
[1] 1
Warning message:
running command 'convert /home/pw/wessanet/rcomp/tmp/2ny3n1500576137.ps
/home/pw/wessanet/rcomp/tmp/2ny3n1500576137.png' had status 1
>
> proc.time()
   user  system elapsed
  0.656   0.076   0.767
```