Application of liquid chromatography-mass spectrometry and chemometrics in the automated characterization of molecular lipid species

Ying-Xu Zeng

Dissertation for the degree Philosophiae Doctor (PhD)

at the University of Bergen

2015

Dissertation date: 05.11.2015

Scientific environment

This dissertation is submitted by Ying-Xu Zeng to the Department of Chemistry, University of Bergen, for the degree of Philosophiae Doctor (PhD). The work was carried out during 2010-2015 at Department of Chemistry, University of Bergen, in cooperation with The Institute of Marine Research (IMR) and The National Institute of Nutrition and Seafood Research (NIFES). The samples were obtained from Haukeland University Hospital, NIFES and IMR. The liquid chromatography mass spectrometry (LC-MS) analyses were mainly carried out at IMR and NIFES. In addition, some high-resolution MS and LC-MS data were acquired at The Swiss Federal Institute of Technology in Lausanne (EPFL), Switzerland, in March 2013.

Acknowledgements

The present work is based on the collaboration with The Institute of Marine Research (IMR) in Bergen, The National Institute of Nutrition and Seafood Research (NIFES) in Bergen, and The Swiss Federal Institute of Technology in Lausanne (EPFL) in Switzerland. I would like to take this opportunity to thank all these institutes for providing excellent research facilities.

I wish to express my sincere gratitude to my supervisor, Prof. Svein Are Mjøs, who provided continuous and immeasurable support, throughout my PhD study. I have benefited greatly from his valuable guidance and expertise in lipids and analytical chemistry.

I am heartily grateful to my supervisor, Prof. Bjørn Grung, for his professional guidance and teaching, particularly in respect of Chemometrics. His great support has been essential during my PhD study.

I am truly indebted to Prof. Pedro Araujo, for his dedicated guidance and great support in the triacylglycerol analysis project performed at NIFES. His insightful advices and warm encouragement have provided me a source of inspiration.

I owe appreciation to Dr. Sonnich Meier, who has provided me great support for the LC-MS analysis at IMR and donated phospholipid standards. Warm thanks also go to technicians at IMR who helped me solving the experimental issues related to LC-MS analysis.

I would like to acknowledge Dr. Zhen-Yu Du, for donating the mouse samples and for offering the opportunity of collaboration. His contributions and advices in the manuscript are truly appreciated.

Sincere gratitude to Dr. Fabrice David from EPFL for providing me great opportunity of collaboration and a warm research stay at EPFL. Kind thanks also go to other personnel in EPFL who assisted me in the high-resolution LC-MS analysis and practical issues.

Many thanks to Haukeland University Hospital and NIFES for providing the mouse brain samples, and thank IMR for providing cod brain samples.

Further warm thanks to my colleagues and friends, for your friendship and support all these years. Thank the administrative staffs at UiB for all the help and assistance during my study here.

Finally, I would like to thank my family, especially my husband, for their boundless love and support and for creating a great positive and joyful atmosphere at home, which is crucial for my PhD study. Special thanks go to my son for being healthy and lovely, and for being a source of happiness.

Bergen, May 2015

Abstract

Lipidomics is an important field that has attracted extensive interest worldwide, due to the increasing awareness of crucial lipid functions in biological systems. Lipidomics aims at detecting, characterizing and quantifying lipid species comprehensively. In the work for the present thesis, analytical strategies based on liquid chromatography-mass spectrometry (LC-MS) and chemometrics were developed for characterization of molecular species of major lipid classes, *i.e.* triacylglycerols (TAG) and glycerophospholipids (GPL) from marine oils and biological systems.

The applicability of liquid chromatography electrospray tandem mass spectrometry (LC-ESI-MS^2) for the structural characterization of naturally occurring TAG in cod liver oil was investigated. A computational algorithm was developed to automatically interpret mass spectra and elucidate TAG structures, and the results of the algorithm were compared against the lipase benchmark method. It was proved that LC-ESI-MS^2 provides a suitable and powerful strategy for the structural characterization of TAG in cod liver oil.

The thesis also evaluates different strategies for differentiating marine oils by means of principal component analysis (PCA). The TAG composition and four different types of data, including total ion current (TIC) and total mass spectral (TMS) profiles derived from LC-ESI-MS and LC-ESI-MS^2, were used as the datasets for PCA. The results show that using the tandem TMS profiles from LC-ESI-MS^2 experiments was the most rapid and convenient approach for the differentiation of the various marine and plant oils investigated, and for the representation of the characteristic TAG patterns.

The thesis proposes a least square spectral resolution (LSSR) approach for the automated characterization and deconvolution of the main GPL species, *i.e.*, phosphatidylcholine (PC) and phosphatidylethanolamine (PE) in biological extracts. Class-specific scanning methods, such as precursor ion scanning and neutral loss scanning, in LC-MS were applied to acquire the lipidomic dataset. The methodology is based on least squares resolution of spectra and chromatograms from theoretically calculated mass spectra with the isotope distribution. The described algorithm was able to resolve PC and PE species of reference mixtures, porcine brain sphingomyeline, cod and mouse brain lipid extracts.

Recent advances in high-resolution mass spectrometry have revolutionized the lipidomics field by providing high-resolution data. The LSSR methodology was further extended to be compatible with this type of data for an accurate identification and quantification of lipid species. The methodology has been expanded to cover the analysis of other major lipid classes such as GPL, sphingolipids, glycerolipids. Examples for the analysis of natural lipids extracts from egg, porcine brain and bovine liver are presented. The flexibility of the methodology allows supporting more lipid classes and more data interpretation functions, which in turn makes LSSR a promising tool for lipidomic data analysis.

LSSR methodology was applied on LC-MS data to evaluate the effects of methylmercury (MeHg) and EPA on intact PC and PE species in mouse brain. The effects of EPA and MeHg on PC and PE composition in brain were evaluated by PCA and ANOVA. The results demonstrate that EPA reduces the levels of arachidonic acid (AA) containing PC and PE species in brain, while MeHg tends to elevate the levels of AA containing PC and PE species. EPA also significantly increases the levels of *n*-3 polyunsaturated fatty acids (PUFA) containing PC and PE species in brain. The results indicate that EPA may counteract the alterations of the PC and PE pattern induced by MeHg, and thus alleviate MeHg neurotoxicity in mouse brain through the inhibition of AA-derived pro-inflammatory factors.

The LSSR methodology was further applied to evaluate the effects of MeHg and EPA on the PC and PE composition in mouse liver and plasma by PCA and ANOVA in conjunction with biological and toxicological analyses. Similar to results from brain, EPA significantly elevates the levels of PC and PE species that contain *n*-3 PUFA and reduces the levels of PC and PE species that contains AA. MeHg increases the levels of PC and PE species with AA to a lower extent. MeHg induces more prostaglandin E2 and less prostaglandin E3, thus increasing pro-inflammatory factors, while EPA displays the ability to decrease the AA-derived inflammatory factors. The histological analysis of cell damage and necrosis and the measurements of biochemical indexes also indicate that MeHg induced chronic inflammatory symptoms in mice, and that EPA can alleviate the MeHg-induced hepatic toxicity. Collectively, EPA may have protective effects against MeHg-induced toxicity in mice due to the favourable modification of membrane phospholipid composition and the inhibition of inflammatory factors release.

In summary, the described strategies and algorithms represent promising tools for the analysis of TAG and GPL species in oils, fats and biological systems. The application of these

methodologies on different objects can provide insights into various research areas, such as food and nutrition, health, pharmacology and toxicology.

Contents

List of publications

This thesis is based on the following list of papers.

I. **Elucidation of triacylglycerols in cod liver oil by liquid chromatography electrospray tandem ion trap mass spectrometry**
Y.X. Zeng, P. Araujo, Z.Y. Du, T.T. Nguyen, L. Frøyland, B. Grung,
Talanta 82: 1261–1270 (2010).

II. **Evaluation of different fingerprinting strategies for discriminating marine oils by liquid chromatography iontrap mass spectrometry and chemometrics**
Y.X. Zeng, P. Araujo, B. Grung, L.X. Zhang
Analyst 136:1507–1514(2011).

III. **Least squares spectral resolution of liquid chromatography–mass spectrometry data of glycerophospholipids**
Y.X. Zeng, S.A. Mjøs, S. Meier, C.C. Lin, R. Vadla
Journal of Chromatography A 1280: 23–34 (2013).

IV. **Extension of least squares spectral resolution algorithm to high-resolution lipidomics data**
Y.X. Zeng, S.A. Mjøs, F. David, A. Schmid
Manuscript

V. **Methylmercury increases and eicosapentaenoic acid decreases the relative amounts of arachidonic acid containing phosphatidylcholines and phosphatidyl-ethanolamines in mouse brain**
Y.X. Zeng, Z.Y. Du, S.A. Mjøs, B. Grung, L. K. Midtbø
Lipids (In revision)

VI. **Eicosapentaenoic acid modulates methylmercury toxicity through favorably modifying membrane phospholipids composition and inhibiting inflammatory factors release**
Y.X. Zeng, S.A. Mjøs, P. Araujo, B. Grung, L. K. Midtbø, Z.Y. Du,
Manuscript

Abbreviations

ALA	α-linolenic acid (18:3 *n*-3)
AA	Arachidonic acid (20:4 *n*-6)
APCI	Atmospheric pressure chemical ionization
APPI	Atmospheric pressure photoionization
CN	Carbon number
CODA	Component detection algorithm
CID	Collision induced decomposition
DAG	Diacylglycerols
Da	Dalton
DHA	Docosahexaenoic acid (22:6 *n*-3)
DPA	Docosapentaenoic acid (22:5 *n*-3)
ECN	Equivalent carbon number
EI	Electron impact
EPA	Eicosapentaenoic acid (20:5 *n*-3)
ESI	Electrospray ionization
FFA	Free fatty acids
FT-ICR	Fourier transform ion cyclotron resonance
GC	Gas chromatography
GC-MS	Gas chromatography-mass spectrometry
GPL	Glycerophospholipids
HO-SPC	Higher order-sequential paired covariance
HPLC	High performance liquid chromatography
LA	Linoleic acid (18:2 *n*-6)
LC	Liquid chromatography
LC-MS	Liquid chromatography-mass spectrometry
LC-ESI-MS	Liquid chromatography electrospray single mass spectrometry
LC-ESI-MS^2	Liquid chromatography electrospray tandem mass spectrometry
LSSR	Least squares spectral resolution
LTB_4	Leukotriene B4
LTB_5	Leukotriene B5
LTQ	Linear trap quadrupole
IUBMB	International Union of Biochemistry and Molecular Biology
IUPAC	International Union of Pure and Applied Chemistry
MAG	Monoacylglycerols
MALDI	Matrix-assisted laser desorption ionization

MCQ	Mass chromatographic quality
MeHg	Methylmercury
MS	Mass spectrometry
m/z	Mass-to-charge ratio
NLS	Neutral loss scan
NMR	Nuclear magnetic resonance spectrometry
NPLC	Normal phase liquid chromatography
RPLC	Reversed phase liquid chromatography
PA	Phosphatidic acids
PC	Phosphatidylcholines
PCA	Principal component analysis
PE	Phosphatidylethanolamines
PG	Phosphatidylglycerols
PGE_2	Prostaglandin E2
PGE_3	Prostaglandin E3
PI	Phosphatidylinositols
PLA_2	Phospholipase A_2
PIS	Precursor ion scan
PLS	Partial least squares
PS	Phosphatidylserines
PUFA	Polyunsaturated fatty acids
QqQ	Triple quadrupole
Q-TOF	Quadrupole time-of-flight
SFA	Saturated fatty acids
SIMCA	Soft independent modeling of class analogy
sn	Stereospecific numbering
S/N	Signal to noise ratio
SPC	Sequential paired covariance
SVD	Singular value decomposition
TAG	Triacylglycerols
TIC	Total ion chromatogram
TMS	Total mass spectral
TOF	Time of flight
TXA_2	Thromboxane A2
TXA_3	Thromboxane A3
WMSM	Windowed mass selection method

1 Introduction: Lipids and lipidomics

1.1 Lipids

The term “Lipids” was initially defined as organic compounds that are soluble in organic solvents but insoluble in water, and that are commonly present in or derived from the living organisms [1]. This chemical class covers a broad range of molecules, such as fatty acids, triacylglycerols (TAG), phospholipids, sterols, sphingolipids and terpenes. However, this definition does not cover all lipids in biological systems, since several new classes (such as lipopolysaccharides) are now widely regarded as lipids, even though they are not soluble in organic solvents [2]. The International Lipid Classification and Nomenclature Committee on the initiative of the LIPID MAPS Consortium broadly define lipids as “hydrophobic or amphipathic small molecules that may originate entirely or in part by carbanion based condensations of ketoacyl thioesters (fatty acyls, glycerolipids, glycerophospholipids, sphingolipids, saccharolipids, and polyketides) and/or by carbocation-based condensations of isoprene units (prenol lipids and sterol lipids)” (Figure 1 [3]).

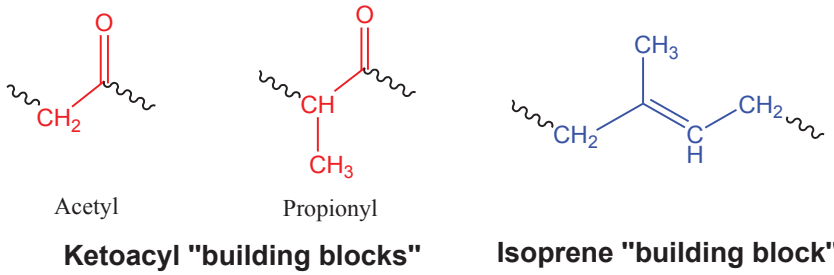


Figure 1 Lipid building blocks. The LIPID MAPS classification system is based on the concept of two fundamental biosynthetic “building blocks”: ketoacyl groups and isoprene groups.

1.1.1 Lipid structures and classification

Lipids are characterized by extreme structural diversity and complexity, with over 37,000 unique structures currently stored in LIPID MAPS, the most comprehensive lipid structure database [4], and up to 180,000 different structures of lipids in theory [5].

Table 1 Lipid categories according to the LIPID MAPS lipid-classification system and the number of structures in LIPID MAPS database.

Categories of lipids	No. of structures in database
Fatty acyls	6954
Glycerolipids	7542
Glycerophospholipids	9387
Sphingolipids	4352
Sterol lipids	2833
Prenol lipids	1257
Saccharolipids	1293
Polyketides	6742

Several sources (such as 'The Lipid Library' [6] and 'Cyberlipids' [7]) propose a simplified classification system based on the number of products upon hydrolysis. Simple lipids (usually neutral) are defined as those yielding at most two types of primary products per mole upon hydrolysis (such as acylglycerols, ether acylglycerols, sterols and their esters and waxes); Complex lipids (usually polar) yield three or more primary hydrolysis products per mole (such as phospholipids and glycolipids) [6, 7]. In 2005, The International Lipid Classification and Nomenclature Committee and The LIPID MAPS Consortium established a comprehensive classification system for lipids that covers both eukaryotic and prokaryotic sources. This commonly accepted system classified lipids into eight well-defined categories and each category is further divided into main classes, subclasses and additional levels of classes (Table 1) [3, 8].

1.1.2 Lipid functions

Lipids play diverse unique and important roles in biological systems. They are the central components of the semipermeable cell membranes whose integrity and physical properties are vital for life processes [9]. In general, lipids primarily reside in cellular membranes and eukaryotic cell membranes, which are mainly composed of glycerophospholipids (GPL), sterols, and sphingolipids (Figure 2) [10]. The membrane of an individual eukaryotic cell has a unique lipid composition [11]. The structural diversity of GPL and sphingolipids originates through the variation of the polar head groups and the apolar hydrocarbon chains, while sterols show little structural variation. Depending on the organism, eukaryotic lipidomes (*e.g.* lipid classes, subclasses, and individual molecular species) may contain thousands of

individual lipid species that structurally and chemically regulate cell membranes [11]. A recent study revealed the enormous structural diversity of lipids in human plasma with over 500 detected lipid species [12]. The construction of a cell membrane does not necessarily require many different lipid species, yet numerous distinct lipid species are endogenously synthesized in the body through a complicated lipid-forming enzyme network. The large variation of lipid species and the resulting diverse physicochemical properties of membrane properties reflect the multiple vital functions of lipids carried out at the cellular, tissue, and organismal levels besides their function as essential cellular constituents [5].

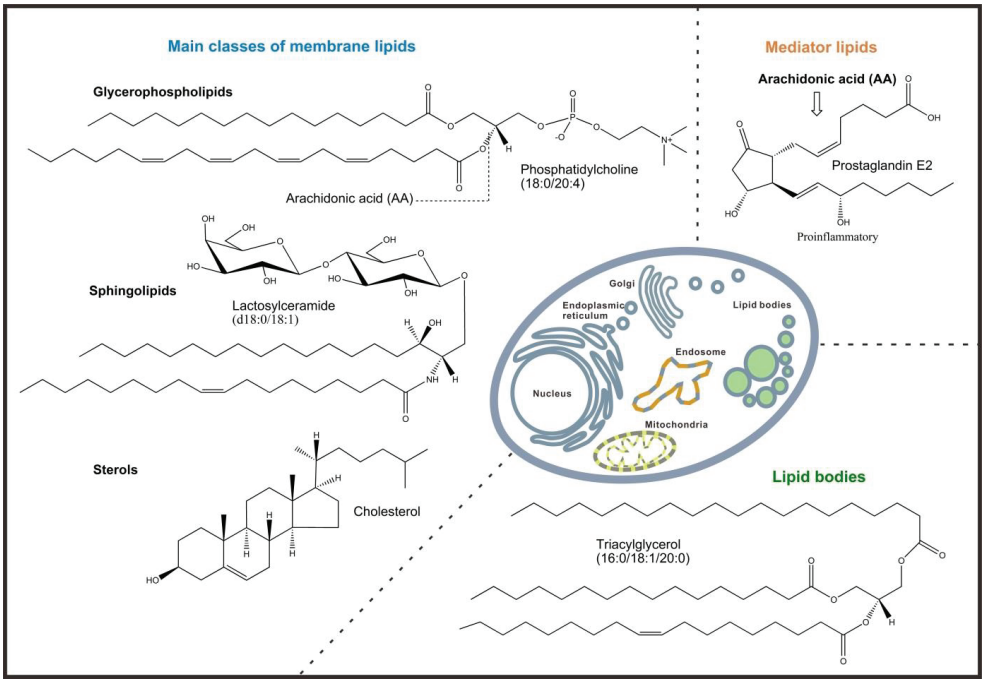


Figure 2 The cellular compartments of common biological lipids. The main classes of eukaryotic cell membranes are composed of GPL, sphingolipids, and sterols (Structures of representative lipids from these classes are shown). Mediator lipids such as eicosanoids can be generated by the metabolism of membrane lipids. Lipid bodies contain nonpolar lipids such as TAG, acting as energy reservoir. Modified from [10].

Studies have indicated that lipids are directly involved in membrane trafficking, cellular signal transduction, energy storage, regulation of membrane proteins, *etc.* [5, 10, 13, 14]. Cell membranes are packed with membrane proteins that can organize the distribution of lipids, which in turn provide an appropriate hydrophobic environment for membrane protein function and interactions [9, 13]. It has been observed that more and more proteins exhibit specific

lipid-binding and/or lipid-interaction capacities, indicating the vital roles of lipid-based membranes involved in protein sorting and signaling [5, 13]. Moreover, emerging bodies of evidence demonstrate that lipid signaling is a vital component of cell signaling. Metabolism of membrane lipids generates highly active mediator lipids, such as eicosanoids, lysolipids, diaylglycerols (DAG), sphingolipids, *etc.*, which control important cellular processes, including cell growth, apoptosis, metabolism and migration (Figure 2). The pathways that generate and respond to signaling lipids have important roles in inflammation, cancer and metabolic syndrome [14]. In addition, some specific lipid species, such as TAG and sterol-esters, function as the energy storage reservoir in living organisms and are stored in the lipid bodies within cells (Figure 2) [10]. On the other hand, the excess of these energy storage species is associated with many health problems, such as obesity and cardiovascular diseases [15, 16].

1.2 Fatty acids

Fatty acids are carboxylic acids with long hydrocarbon chains, which constitute the starting point in lipid structures. The hydrocarbon chain varies in length, and can be saturated, monounsaturated, or polyunsaturated. The most common and important fatty acids from plant and animal origins contain unbranched hydrocarbon chains with an even number of carbons ranging from 12 to 22. They are essential for energetic, metabolic, and structural functions of all organisms.

1.2.1 Fatty acid nomenclature

Different systems of nomenclature are used for fatty acids.

1.2.1.1 Trivial names

Trivial names (or common names) are non-systematic historical names which are commonly used, for example, palmitic, stearic, or oleic acids, as shown in Table 2.

1.2.1.2 Systematic nomenclature

Systematic nomenclature defined by standard International Union of Pure and Applied Chemistry (IUPAC) names a fatty acid after the length of its parent hydrocarbon chain (Table 2). Double bonds are designated by counting from the carboxylic acid end (α end) and are labelled with *cis*-/*trans*- notation or *E*-/*Z*- notation (Figure 3). For example, oleic acid is *cis*-9-

octadecenoic acid (or Δ9-octadecenoic acid), a carboxylic acid (oic) with 18 carbon atoms (octadec) and one olefinic centre (en) which lies between carbon 9 and 10 (counting from the carboxyl end) and has *cis* configuration (Figure 3).

Table 2 Terms and symbols used for designating major fatty acids [17].

Trivial name	*n*-x system	Δ system	Systematic name (IUPAC)
Myristic	14:0	14:0	*n*-Tetradecanoic acid
Palmitic	16:0	16:0	*n*-Hexadecanoic acid
Palmitoleic	16:1 *n*-7	Δ9c-16:1	*cis*-9-Hexadecanoic acid
Stearic	18:0	18:0	*n*-Octadecanoic acid
Oleic	18:1 *n*-9	Δ9c-18:1	*cis*-9-Octadecanoic acid
Linoleic	18:2 *n*-6	Δ9c,12c-18:2	*cis,cis*-9,12-Octadecadienoic acid
α-Linolenic	18:3 *n*-3	Δ9c,12c,15c-18:3	all-*cis*-9,12,15-Octadecatrienoic acid
γ-Linolenic	18:3 *n*-6	Δ6c,9c,12c-18:3	all-*cis*-6,9,12-Octadecatrienoic acid
Gadoleic	20:1 *n*-9	Δ11c-20:1	*cis*-9-Eicosenoic acid
Arachidonic	20:4 *n*-6	Δ5c,8c,11c,14c-20:4	all-*cis*-5,8,11,14-Eicosatetraenoic acid
EPA	20:5 *n*-3	Δ5c,8c,11c,14c,17c-20:5	all-*cis*-5,8,11,14,17-Eicosapentaenoic acid
DHA	22:6 *n*-3	Δ4c,7c,10c,13c,16c,19c-22:6	all-*cis*-4,7,10,13,16,19-Docosahexaenoic acid
Erucic	22:1 *n*-9	Δ13c-22:1	*cis*-13-Docosaenoic acid

1.2.1.3 The *n*- or ω-reference nomenclature

Shorthand nomenclature of fatty acids (*n*- or ω) is also in common usage, and is composed of the carbon number in the fatty acid chain followed by the number of double bonds (Table 2). For example, saturated fatty acids (SFA) such as stearic acid is denoted as '18:0' or 'C18:0' (Figure 3). The position of the double bonds can be specified by counting from the methyl end of the fatty acid (ω end). For example, *n*-3 and *n*-6 (or ω3 and ω6) denote fatty acids with the first double bond at the third and sixth carbon, respectively, counting from the methyl end (ω end). Linoleic acid (LA) can be abbreviated as '18:2 *n*-6, 'C18:2 *n*-6', or '18:2ω6' (Figure 3) [17].

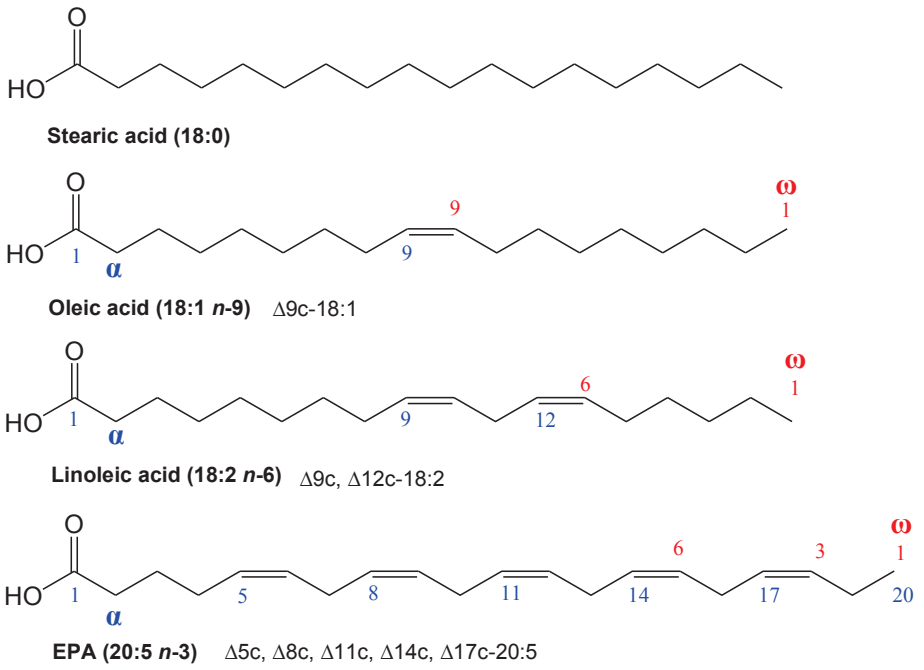


Figure 3 Representative fatty acids structures and nomenclature.

1.2.1.4 Carboxyl-reference nomenclature (Δ)

In this system, each double bond is indicated by Δx, counting from the carboxylic acid end (α end). In addition, each double bond is designated by a *cis*- or *trans*- prefix, or the abbreviated letters '*c*' or '*t*', indicating the conformation of the molecule around the bond. For example, the polyunsaturated fatty acid (PUFA) such as eicosapentaenoic acid (EPA, 20:5 *n*-3) can be named as all *cis*-5,8,11,14,17-20:5, or Δ5c, Δ8c, Δ11c, Δ14c, Δ17c-20:5 (Figure 3).

1.2.2 Polyunsaturated fatty acids

1.2.2.1 Biological properties of polyunsaturated fatty acids

PUFA are fatty acids of 18 or more carbons in length with two or more double bonds, and they can be classified into two major groups, the *n*-6 and *n*-3 families. Linoleic acid (LA, 18:2 *n*-6), a long-chain *n*-6 PUFA, and α-linolenic acid (ALA, 18:3 *n*-3), a long-chain *n*-3 PUFA, are essential fatty acids that must be obtained through diet since the human body cannot synthesize them de novo. In addition, the human body cannot convert *n*-6 into *n*-3 PUFA since it lacks enzymes for forming double bonds (desaturase enzymes) past the Δ9 position.

The other PUFA can be synthesized through a common desaturase/elongase system, LA can be elongated and desaturated to arachidonic acid (AA, 20:4 *n*-6), and ALA can be elongated and desaturated into EPA and then into docosahexaenonic acid (DHA, 22:6 *n*-3) (Figure 4) [18]. The position of the double bond from the methyl end never changes during physiological transformations in the human body. Studies have shown that the conversion of dietary ALA to EPA is low and conversion to DHA is further limited. The whole-body conversion of ALA to DHA is < 5% in humans and depends on the concentration of long chain PUFAs in the diet [19]. The conversion process can be further depressed by various factors such as aging and disease [20].

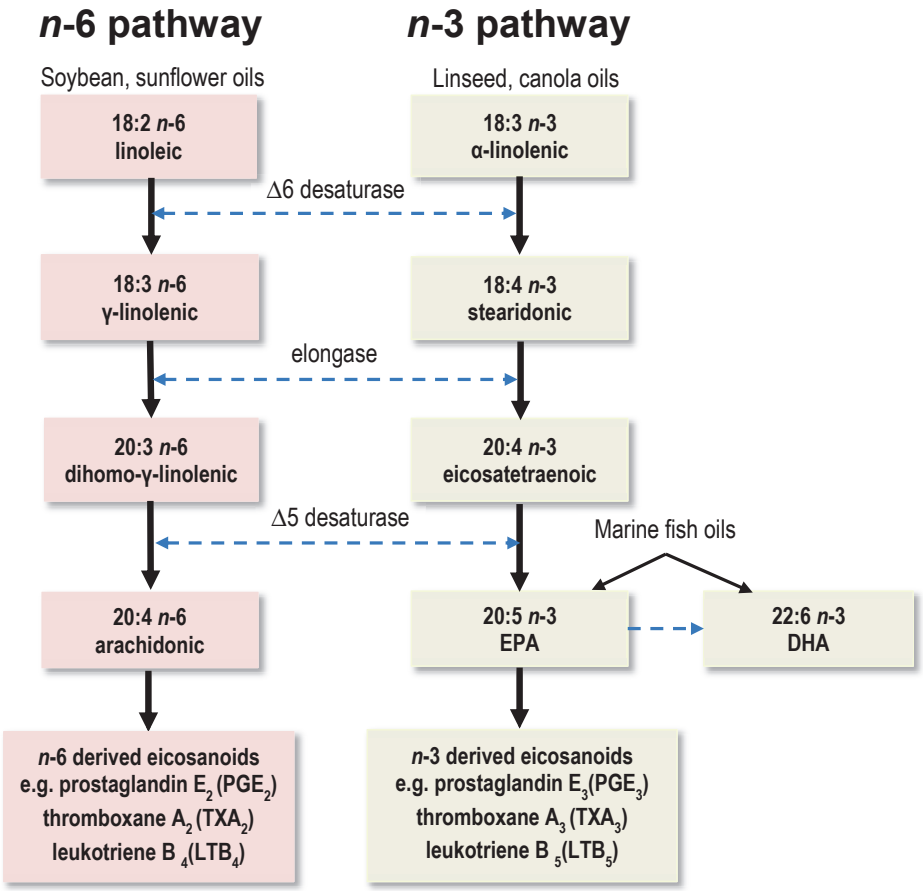


Figure 4 Metabolic pathways of the *n*-3 and *n*-6 PUFAs and eicosanoids production in mammals. Modified from [18].

In particular, the metabolism of *n*-3 and *n*-6 PUFA is of great interest because of the biological actions of their in vivo metabolites (eicosanoids). Both AA and EPA are precursors of eicosanoids. AA derived eicosanoids, such as prostaglandin E2 (PGE_2), leukotriene B4 (LTB_4) and thromboxane A2 (TXA_2), exhibit pro-inflammatory effects, while EPA derived eicosanoids, such as prostaglandin E3 (PGE_3), leukotriene B5 (LTB_5) and thromboxane A3 (TXA_3), tend to inhibit platelet aggregation and display anti-inflammatory effects [21, 22]. Recent studies have shown that these bioactive eicosanoids play a key role in the inflammatory process, which is closely associated with chronic diseases such as rheumatoid arthritis, asthma, atherosclerosis, obesity, Crohn's disease, cancer, *etc*. [23, 24]. A detailed description of the biosynthesis of PUFA by elongation and desaturation of the carbon chain is shown in Figure 4 [18].

1.2.2.2 Importance of polyunsaturated fatty acids

PUFA are important structural components that regulate membrane fluidity and selective permeability. For example, DHA and AA are very abundant in neuronal tissues such as brain and retina [25, 26]. Deficiencies in both DHA and AA have been associated with disorders of the neuro-visual development and other complications of premature birth [27, 28]. In addition, PUFA serve as precursors for eicosanoids, growth regulators and hormones, and they are constituents of membrane phospholipids involved in signal transduction [29, 30].

The first epidemiologic studies conducted in the 1970s, indicated that *n*-3 PUFAs, such as EPA and DHA, are important dietary components for health and disease prevention [31, 32]. Since then, numerous observational studies, randomized controlled trials, and clinical, animal, and in vitro studies suggest that increased intake of *n*-3 PUFAs from fish or fish-oil supplements reduces the risk of cardiovascular disease and reduces all-cause mortality, sudden death, and stroke [33-35]. The health effects of *n*-3 PUFAs, such as EPA and DHA, have been extended to beneficial effects on Crohn disease, asthma and chronic obstructive pulmonary diseases, alleviation of symptoms of cystic fibrosis, cancers of the breast, colon, and prostate, prevention of inflammatory and autoimmune disorders (rheumatoid arthritis, psoriasis) and improvement in growth and development [36-39]. The recommended intake of *n*-6/*n*-3 PUFA ratio is between 1/1 and 4/1 [40], while the present dietary pattern indicates a much higher ratio of *n*-6/*n*-3 PUFA (15/1–16.7/1) [41].

1.3 Triacylglycerols

1.3.1 Structure and composition

Triacylglycerols (TAG), which are made up of three fatty acid molecules esterified to a glycerol backbone, are the main components of most of the dietary oils and fats from plant and animal origins (> 98 %). The molecular structure of each individual TAG species can be described basically by three main attributes [42]:

– the total carbon number (CN) defined as the sum of the alkyl chain lengths of the three fatty acids,
– the degree of unsaturation in each fatty acid, and
– the position and configuration of the double bonds in each fatty acid.

An example of a typical TAG molecule is shown in Figure 5. TAG species can differ in the length of fatty acyls, the number of double bonds in the fatty acyl chain and the position of double bonds in the fatty acyl chain. These variations will generate a large number of TAG molecules.

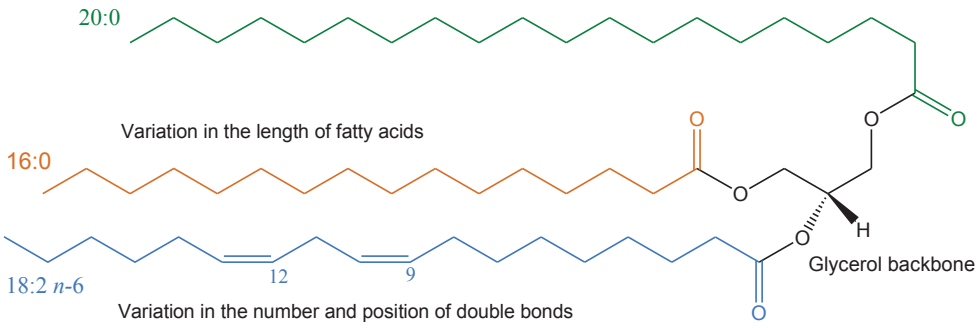


Figure 5 Example of a TAG molecule structure (1- hexadecanoyl-2-(9*Z,* 12Z-octadecadienoyl)-3-*sn*-eicosanoy-glycerol). The glycerol backbone is indicated in black. Three fatty acids (green, orange, blue) are esterified to glycerol.

Moreover, each TAG species may be differentiated in regiospecific/stereospecific isomers by the positioning of the three fatty acids on the glycerol backbone, since the trihydric alcohol glycerol itself has a plane of symmetry. When the two primary hydroxyl groups are esterified with different fatty acids, the resulting TAG can be asymmetric and can therefore display

optical activity. The stereochemistry of TAG can be represented by a Fischer projection and the "stereospecific numbering" (*sn*) system as recommended by IUPAC and the International Union of Biochemistry and Molecular Biology (IUBMB) commission on the nomenclature of glycerolipids [43].

A Fischer projection of a natural *L*-glycerol derivative is shown in Figure 6. The secondary hydroxyl group is labelled as position *sn*-2. The carbon atom above this then becomes *sn*-1 position while the carbon below becomes *sn*-3 position. A single molecular species is identified by listing the *sn*-1, *sn*-2 and *sn*-3 positions in this particular order [42].

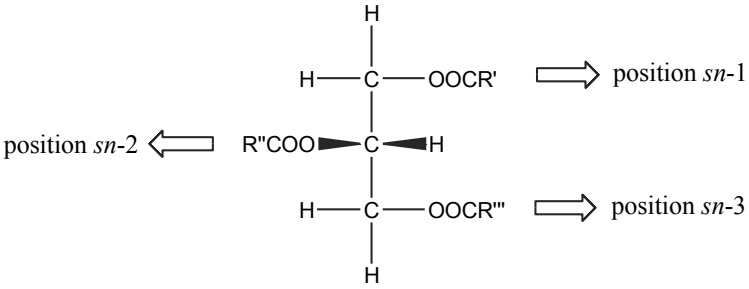


Figure 6 Schematic structure of a Fischer projection of a TAG molecule.

The potential number of TAG is quite large and it rises very quickly with the number of fatty acids present in the pool (Table 3). For example, a fat containing only 2 different fatty acids results in the theoretical number of 8 possible TAG; or 6 TAG if stereoisomers are excluded and 4 TAG if all isomers are excluded.

AAA	AAB	ABA	BAA	ABB	BAB	BBA	BBB
A_3	A_2B			AB_2			B_3

Vegetable oils typically contain 5-10 abundant fatty acids, which may give 125-1000 individual TAG molecules in theory (Table 3). However, the situation is even more complex with samples such as oils derived from fish or marine mammals that typically contain 20-40 fatty acids in significant amounts [44].

Table 3 Relation between number of fatty acids (N) and number of TAG

Fatty acids	Number of TAG		
N	All TAG N^3	Excluding stereoisomers $(N^3+N^2)/2$	Excluding all isomers $(N^3+ 3N^2+2N)/6$
2	8	6	4
3	27	18	10
4	64	48	20
5	125	75	35
10	1000	550	220
20	8000	4200	1540
40	64000	32800	11480

1.3.2 Digestion, absorption and metabolism

The digestion, absorption, and metabolism of TAG are efficient and relatively well-defined processes. The first step in the digestion of TAG, which takes place in the stomach, is a partial enzymatic hydrolysis into diacylglycerols (DAG) and free fatty acids (FFA), performed by lingual lipase and possibly gastric lipase [45, 46]. Both lipases preferentially hydrolyze the *sn*-3 ester bond resulting in formation of *sn*-1,2-DAG [45, 47]. Approximately 30 % of total dietary TAG may be digested in the stomach [48]. The products remaining in the stomach after hydrolysis, *i.e.*, FFA, DAG, and monoacylglycerols (MAG), contained in emulsion droplets, are propelled through the pylorus into the duodenum (Figure 7).

The major digestion of TAG results from hydrolysis with pancreatic lipase in the intestine. Pancreatic lipase acts in conjunction with co-lipase and bile salts to digest TAG (Figure 7). The process of hydrolysis is regiospecific since pancreatic lipase preferentially hydrolyzes fatty acids from *sn*-1 or *sn*-3 positions of the TAG, with the release of *sn*-2-MAG and FFA [49]. Isomerization of the *sn*-2-MAG to *sn*-1 or *sn*-3-MAG occurs to some extent, and these can be degraded completely to glycerol and FFA [47]. In the human adult, most of the fatty acids in the *sn*-2 position remain intact as *sn*-2-MAG during digestion and absorption since the rate of hydrolysis at the *sn*-2 position of the TAG is very slow [49]. The lipolysis products including FFA, MAG and DAG are solubilised together with phospholipids and cholesterol by lysophospholipids and bile salts into micelles and thus absorbed (Figure 7). Research also indicates better absorption of SFA as *sn*-2-MAG rather than as FFA [49-51]. Within the intestine, the fatty acids in the *sn*-2 position of the MAG and the fatty acids released from the *sn*-1 or *sn*-3 position of the TAG are absorbed in mucosal cells and re-synthesized into TAG, thereby the fatty acids at the *sn*-2 position in dietary TAG are conserved.

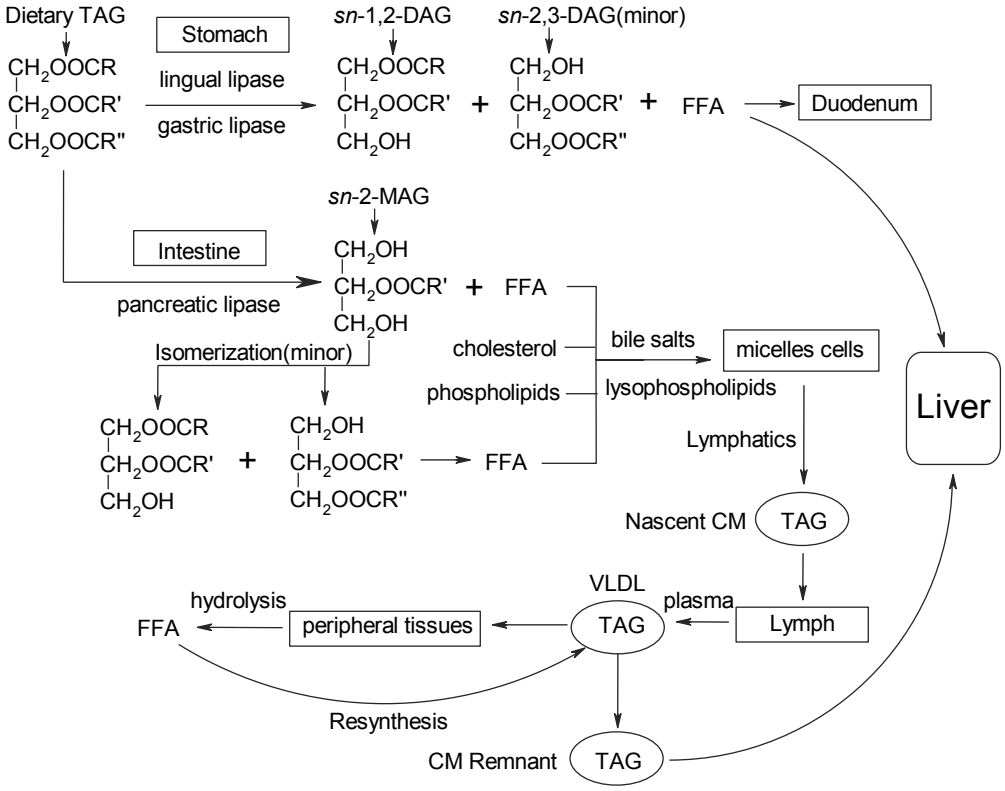


Figure 7 A representative scheme of the digestion, absorption and metabolism of TAG.

Generally, fatty acid chain length and degree of unsaturation as well as the positional distribution of fatty acids in dietary TAG profoundly affect digestion, absorption, and metabolism of dietary fats [52, 53]. It has been indicated that the composition of position *sn*-2 is of great importance when TAG are consumed and digested by the body, probably because *sn*-2-MAG are formed and subsequently absorbed by the intestine [49]. Christensen *et al.* have reported that EPA and DHA were more efficiently absorbed in rats when they were located in the *sn*-2 position of TAG than when they were randomly distributed among the three positions [54].

1.4 Phospholipids

1.4.1 Structure and composition

Phospholipids are the major classes of cell membrane lipids that form the lipid bilayers. The predominant phospholipid species in the eukaryotic cells are glycerol-based lipids referred to as GPL, accounting for approximately 60 % of the lipid mass [9]. GPL are derived mainly from *sn*-1,2-DAG with a phosphate residue in *sn*-3 position that is linked to a simple organic molecule called the head group (Figure 8).

GPL can be grouped into different classes depending on the identity of the head group. The major classes of GPL found in mammalian cell membrane include phosphatidylcholines (PC), phosphatidylethanolamines (PE), phosphatidylserines (PS), phosphatidylglycerols (PG), phosphatidylinositols (PI), phosphatidic acids (PA), as shown in Figure 8. Among these, PC and PE are the dominant classes in most eukaryotic membranes, and are present in about 3:2 molar ratio and constitute around 75 mol% of total GPL. The other classes of phospholipids constitute around 20 mol% of total GPL [9]. Phospholipids usually distribute asymmetrically in the membranes. For example, PC mainly reside in the outer monolayer, while PE, PS and PI are distributed in the inner monolayer of plasma membranes [55].

In addition, fatty acids at *sn*-1 position can be linked by ether or vinyl ether bonds in some of the classes, corresponding to plasmanyl and plasmenyl GPL species, as shown in Figure 9. These plasmanyl and plasmenyl subclasses are found in minor proportions in cell membranes and they mainly exist in the PC and PE classes [9]. The complexity of GPL species not only lies in the variation of head groups and differences in the linkages of the *sn*-1 hydrocarbon chain, but also lies in the length of the carbon chains, the number of double bonds and the position of the double bonds in the carbon chains, generating an enormous number of GPL species. Investigators have estimated 9,600 phospholipid species (of which ~8,000 GPL are annotated in The LipidMaps Database [4]) based on 40 common fatty acids as building blocks [5].

Another category of phospholipids is sphingolipids, which contain a backbone of a sphingoid base such as sphingosine (Figure 10). This category represents around 5-10 mol% of total lipid mass in most brain cells [9]. Sphingolipids can be classified into ceramides, sphingomyeline, cerebrosides, glucosylceramides, lactosylceramides, sulfatides and other

glucosylceramides. Figure 10 shows the structures of typical ceramide and sphingomyelin species. In general, the sphingosines with 18 carbon backbones are the predominant species in most mammalian sphingolipids. Sphingosine backbones containing 14-22 carbons are also present in some cases.

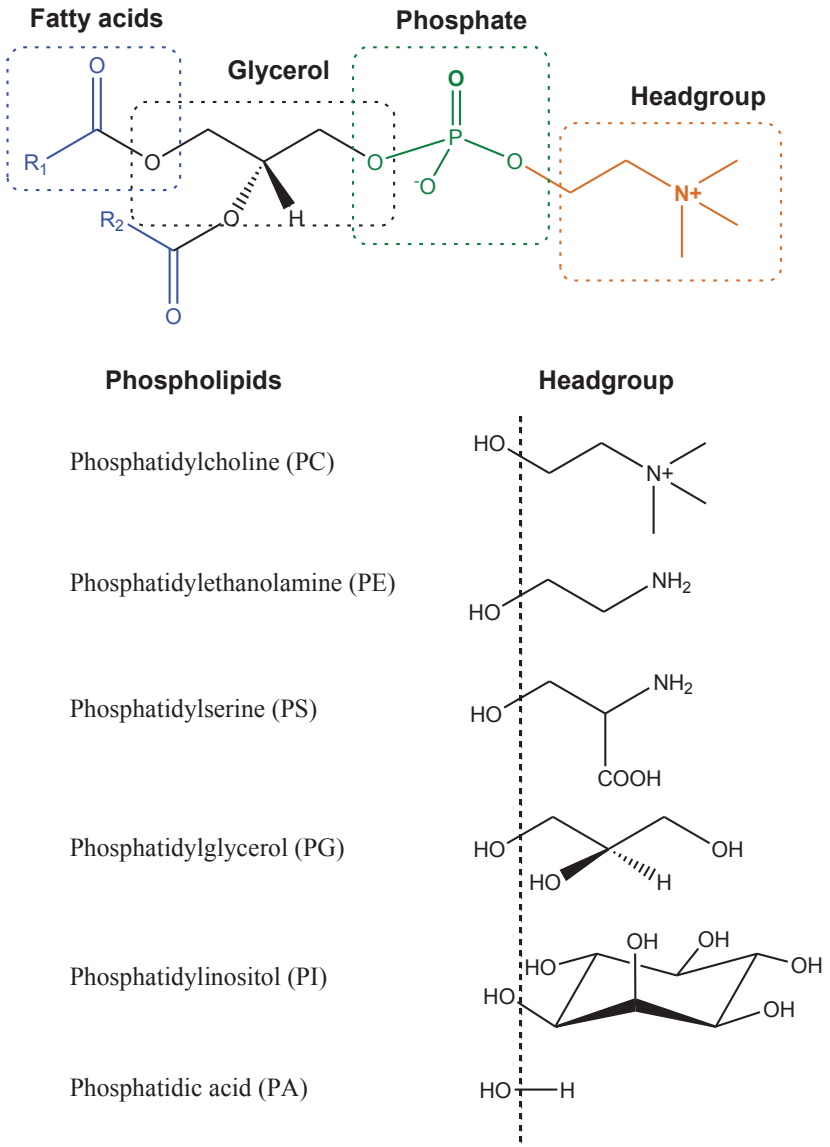


Figure 8 The general structure and classes of phospholipids. GPL consist of a glycerol backbone, two fatty acyl moieties, and a phosphate residue linked to a head group.

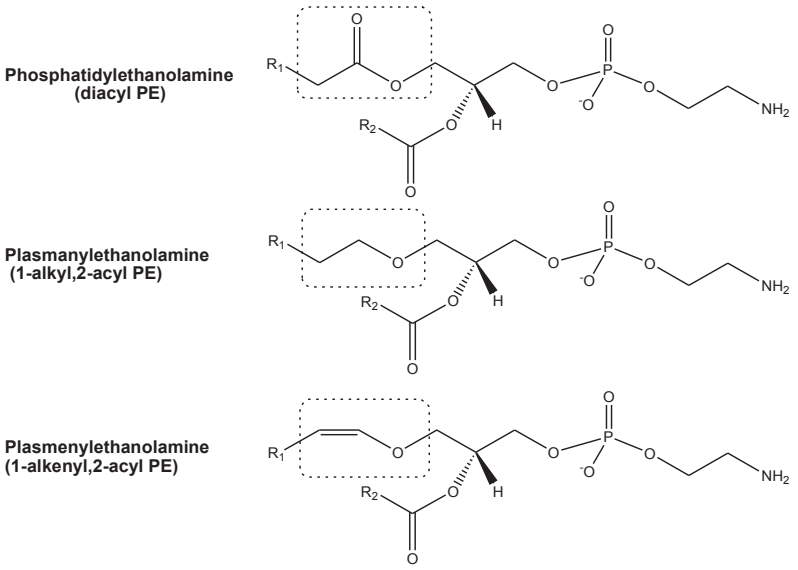


Figure 9 The variation of *sn*-1 linkage of subclasses of ethanolamine phospholipids species. They can be classified as diacyl PE, plasmanyl PE and plasmenyl PE. The other GPL classes can be classified and abbreviated in the similar way.

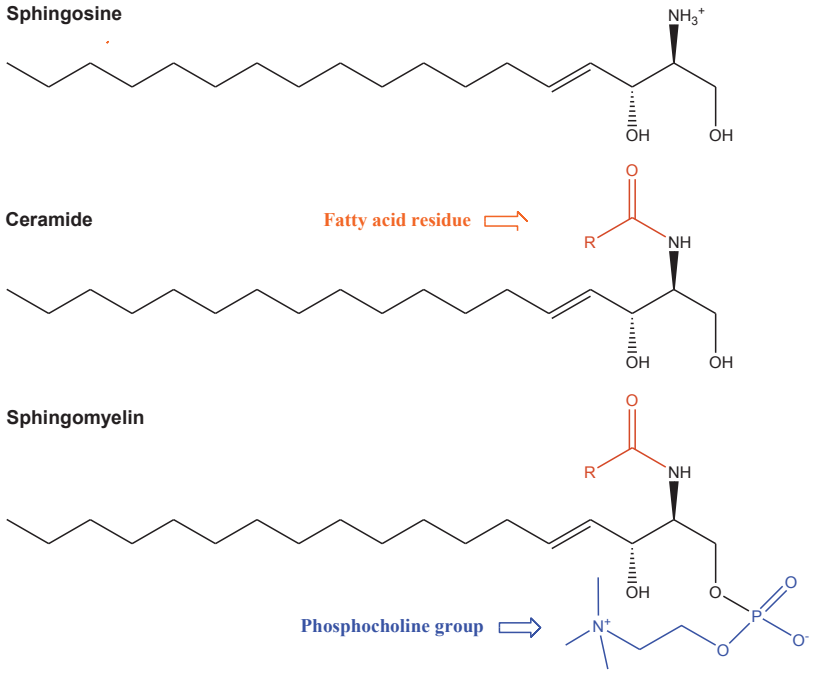


Figure 10 Examples of sphingolipids, ceramide and sphingomyelin species.

1.4.2 Digestion, absorption and metabolism

Intestinal digestion, absorption and metabolism of phospholipids involve dietary and endogenous phospholipids. The dietary phospholipids are mainly derived from eggs, meat, fish and oilseeds, which occur as the components of animal and vegetable cell membranes. The normal dietary intake of phospholipids is 2–8 g per day [56]. The endogenous phospholipids are mainly secreted via the bile and approximately 10–20 g of phospholipids is delivered by the biliary pathway to the intestinal lumen per day [57], which is quantitatively more important than the dietary phospholipids. In particular, PC are the predominant phospholipids from both dietary and endogenous phospholipids sources, which are the second most abundant lipid class in the digestive tract after TAG. PC can be highly absorbed (>90%) by the human intestine, and rapidly occurs in plasma lipoproteins and red blood cells [57]. Since lingual and gastric lipases are incapable of hydrolysing PC, the digestion of phospholipids occurs in the small intestine. The pancreatic phospholipase A_2 (PLA_2) is the main enzyme that hydrolyzes the fatty acid ester bond in the *sn*-2 position of phospholipids and yields products of FFA and lyso-PC [58]. Some other lipases secreted by the pancreas in response to food intake also play a role in the chemical breakdown of phospholipids. The hydrolyzed products are taken up by mucosal cells and are re-secreted to newly-formed phospholipids, deposited into the small intestinal cells at the microvilli [59]. The newly-formed phospholipids join a protein carrier located inside the intestinal cell to form a lipoprotein such as chylomicron. A proportion of chylomicron phospholipids is subsequently transferred to high-density lipoproteins. Deacylation of lyso-PC in the gut lumen by lysophospholipase is believed to be quite limited and the majority of phospholipids is taken up as lyso-PC and FFA [57]. Metabolism of membrane lipids generates highly active signalling lipids such as eicosanoids.

1.5 Lipidomics

With the rapid development of lipid research, the concept of “lipidomics” was introduced in the early 2000s. Lipidomics focuses on the global study of molecular lipids in biological systems and aims at the “full characterization of lipid molecular species and of their biological roles with respect to expression of proteins involved in lipid metabolism and function, including gene regulation” [60]. Advances in mass spectrometry instrumentation and computational methods have greatly promoted the development of the lipidomics field,

enabling precise identification and quantification of the lipidome in cells, tissues, and biofluids. Through the incorporation of multiple lipidomics techniques, lipidomics have provided insight into the functional roles of subcellular membrane compartments in mammalian cells and have expanded our knowledge on human health and diseases [61]. Nowadays, lipidomics have been recognized as an essential tool for the study of many diseases and physiological processes, such as atherosclerosis, Alzheimer's disease, some cancers, and inflammatory processes [62]. The ultimate goal of lipidomics is to understand the role of lipids in the biology of living organisms. This tool can be integrated with multidisciplinary sets of data derived from molecular-profiling techniques such as genomics, transcriptomics, and proteomics to solve various questions of biological importance. Lipidomics approaches can be applied to study the following main research areas.

Food, nutrition and health

Dietary oils and fats, including various animal fats and plant oils, are important constituents of the human diet. In particular, lipids from fish and marine mammals have been widely employed as dietary supplements and food ingredients due to their documented health benefits derived from the significant contents of essential *n*-3 PUFA. The quality and nutritional values of these products largely depend on their lipid composition. Lipidomics can be effectively used to assess the quality and to evaluate the nutritional values of these products. In addition, lipidomics can be used to characterize the effects of specific diets or nutrients in physiological contexts and to elucidate the interactions between diet, nutrients, and lipid metabolism [63, 64].

Health, pathology and disease diagnostics

Epidemiological studies have shown that lipid composition and metabolism reflect health status and are closely associated with many diseases, such as obesity, diabetes, cancer, cardiovascular and neurodegenerative diseases [64-66]. Lipidomics strategies can be used in the biomarkers discovery, disease diagnosis and assessment of drug or nutritional supplementation. There are several branches of lipidomics targeting specific organs or body systems: brain and central nervous system lipidomics, ocular lipidomics, skin lipidomics, lungs and respiratory system lipidomics, circulation and cardiovascular lipidomics, skeletal lipidomics, and cancer lipidomics. More details of the application of lipidomics in the health field can be found in the following reviews [63, 65].

Other fields

Lipidomics can be utilized to study the effects of pharmacological treatments and toxic products (*e.g.*, contaminants, chemicals, drugs of abuse) on lipid metabolism [64].

In addition to the above fields, lipidomics can also be applied in the research fields of cell biology, biochemistry, genetics, transcriptomics, proteomics, biophysics, ecophysiology and organic geochemistry.

2 Lipidomics techniques

In this chapter, the main lipidomics techniques used in the present thesis will be presented. A typical workflow of a lipidomic platform is illustrated in Figure 11, which includes the lipid extraction from different types of samples, mass spectrometry (MS) based analysis, and automated data analysis in conjunction with chemometrics.

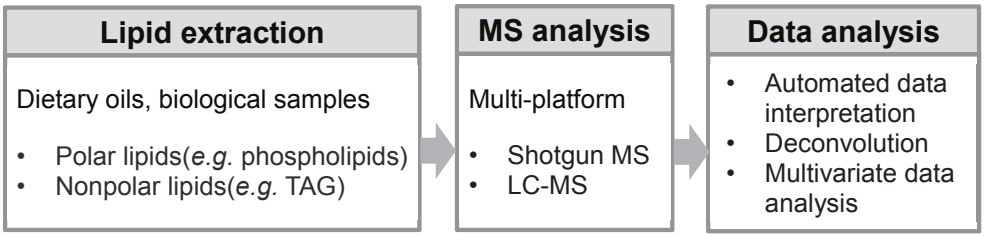


Figure 11 Typical workflow of lipidomic platform. Abbreviation: LC-MS: Liquid Chromatography-Mass Spectrometry

2.1 Fundamentals of lipidomics mass spectrometry

In this section, a brief overview of the main components of a mass spectrometer used in the work for this thesis will be given. The essential components of a mass spectrometer system include an inlet, which introduces the sample in solution, an ion source to ionize the sample, a mass analyzer to separate the ions of compounds based on the mass-to-charge (*m/z*) ratio, and an ion detector (Figure 12) [67, 68].

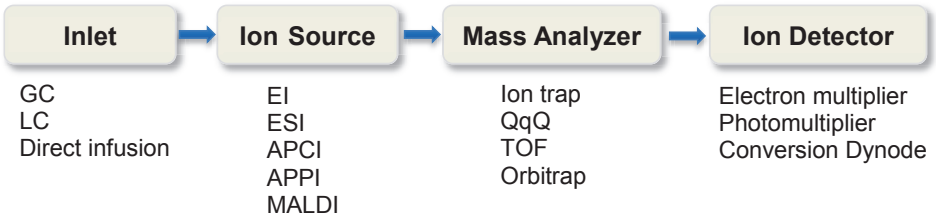


Figure 12 The essential components of a mass spectrometer system and examples of some of the important techniques applied to lipidomics. Abbreviations: EI: Electron Impact; ESI: Electrospray; APCI: Atmospheric Pressure Chemical Ionization; APPI: Atmospheric Pressure Photoionization; MALDI: Matrix-Assisted Laser Desorption Ionization; QqQ: Triple Quadrupole; TOF: Time of Flight. Modified from [68].

The current major lipidomic platform is either based on LC-MS or MS techniques (also referred as shotgun lipidomics). These approaches have their own merits and disadvantages. For example, shotgun lipidomics is very fast and therefore appropriate for high-throughput and large scale analysis of lipids [69]. However, it has a limited sensitivity for the less abundant molecular lipid species, as well as limitations in distinguishing isobaric lipid species. LC-MS lipidomics is quite time consuming as it needs the LC separation of lipid species. However, LC-MS based lipidomics can reduce interferences from the high abundance lipid species and allow the analysis of low abundance lipid species [70]. In addition, the LC separation helps reduce the complexity of samples and increase the accuracy in identification of lipid species.

2.1.1 Reversed phase LC

The implementation of the hyphenated techniques of LC and ESI-MS has become the most widely employed method in lipidomics analyses of various samples [62]. LC techniques can be classified into reversed phase (RP) LC and normal phase (NP) LC. NPLC consists of polar stationary phases and non-polar mobile phases, while RPLC consists of non-polar stationary phases and polar mobile phases. RPLC is the dominating separation technique in lipidomics [71]. The separation mechanism in RPLC depends on the hydrophobic interaction between the solute molecule in the mobile phase and the hydrophobic stationary phase. The molecules that are more hydrophobic will have longer retention times. Decreasing the mobile phase polarity can reduce the hydrophobic interaction between the solute and the stationary phase and promote desorption.

Lipid species are separated according to chain length and degree of unsaturation on RP columns [72, 73]. For the compounds analysed on a RPLC system, the equivalent carbon number (ECN) of the compounds can be calculated by the following equation:

$$ECN = CN - 2DB \qquad \text{Equation 1}$$

where CN is the total number of fatty acid carbons in the molecule and DB is the total number of double bonds in the fatty acids.

A mathematical relationship between the ECN of a lipid species and the relative retention time in the RPLC system has been calculated by Brouwers [74]. The use of elution pattern of lipid species on a RPLC system can be used to assist in the identification of the lipid compounds. For example, the retention of PC species in RPLC system follows their ECN

values. Specifically, the PC species having the same ECN values are basically in the same retention time region and the PC species are gradually eluted with increasing ECN values. This rule can be used to exclude some false identification cases based on the spectra alone [75].

2.1.2 Electrospray ionization

The development of electrospray ionization (ESI) has revolutionized the mass spectrometric analysis of biomolecules, including lipids. In 1994, the first combination of LC and ESI-MS for phospholipids analysis was published [76]. Since then, ESI has become the most frequently employed ionization technique in the lipidomics field. ESI is a soft ionization technique with minimal in-source fragmentation. The basic principle of this technique is shown in Figure 13. In an ESI experiment, the liquid sample is introduced from an LC column or infusion pump to the probe near the orifice of a mass spectrometer. The high potential difference between the probe tip and the sample cone causes electroionization of the liquid sample. Charged droplets are generated with the help of a drying gas. The surface of the droplets that contains the ionized compounds will become charged, either positively or negatively. As the solvent evaporates, the size of the droplets is reduced, and, consequently, the density of charges at the droplet surface increases. The repulsion forces between the charges increase until there is an explosion of the droplet (Coulomb explosion). This process repeats until analyte ions are ejected from the droplet [76, 77].

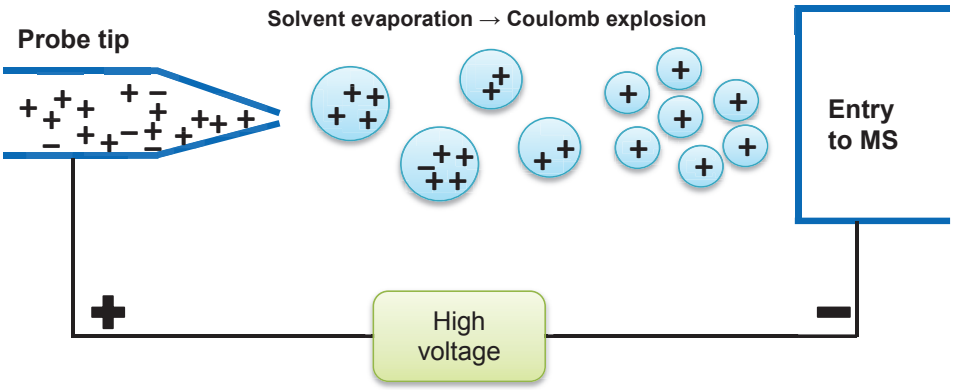


Figure 13 Schematic diagram of the principle of electrospray (ESI) ionization in positive mode. Modified from [78].

The ESI technique allows rapid, accurate and sensitive analysis of a wide range of analytes from low molecular weight (less than 200 Da) polar compounds to biopolymers larger than 100 kDa. Generally, compounds of less than 1000 Da produce singly charged molecules and compounds of high molecular weight produce a series of multiple charged ions. The typical sample flow rates of the ESI ranges from 5 to 30 μL/min. Nanoelectrospray today supports stable flow rates as low as 20 nL/min, which further increases sensitivity [5].

2.1.3 Characterization of lipids by ESI-MS

ESI is a soft ionization technique where the in-source collision induced decomposition (CID) is minimized. In positive-ion mode, positively charged lipid adducts will be formed. Formation of these adducts will be affected by the concentration and affinity of the small cations for each lipid species. Sodium ions which have high affinity for polar lipids such as phospholipids are the most common adducts in the analysis of crude lipid extracts, and are even applied for non-polar lipids, such as TAG [79]. Therefore, sodiated adducts ($[M+Na]^+$) would normally be the most abundant masses in the positive-ion ESI mass spectra if no other modifiers are added to the solution or mobile phase. This is particular the case for polar lipid species [79, 80]. When organic acids (*e.g.*, formic, acetic acid) are employed as modifiers, protonated adducts ($[M+H]^+$) will be generated for the majority of lipid classes. When ammonium salts (*e.g.*, ammonium formate or acetate) are used as modifiers, protonated adducts ($[M+H]^+$) will be formed for phospholipid species in most cases and ammonium adducts ($[M+NH_4]^+$) for TAG species [75, 81]. For example, $[M+H]^+$ but not $[M+NH_4]^+$ were detected for both PC 16:0/18:1 and PE 16:0/18:1 in positive ion mode (Figure 14a, c).

In the negative-ion mode, the lipid species are detected as either deprotonated adducts ($[M-H]^-$) or anionic adducts. If the lipid species carry a separable ionic bond, such as PE, PI, PS, PA, PG, CL, FFA, eicosanoids *etc.*, deprotonated adducts will be generated [9, 79, 82]. If the lipid species belong to polar neutral classes (*e.g.*, cerebroside, glycolipids) or strong zwitterionic lipid classes (*e.g.*, PC and sphingomyelin), anionic adducts (*e.g.*, $[M+CH_3COO]^-$, $[M+Cl]^-$, $[M-CH_3]^-$, and $[M+HCOO]^-$) will be formed [79, 80, 83]. For example, deprotonated adducts ($[M-H]^-$) were detected for PE 16:0/18:1 (Figure 14d), while $[M-CH_3]^-$ and $[M+HCOO]^-$ ions were observed for PC 16:0/18:1 (Figure 14b). In addition, the MS/MS spectra of GPL usually yield fragments of the acyl anion of fatty acids, *i.e.*, $[R_1COO]^-$ and $[R_2COO]^-$, as shown in Figure 14b and d. The characteristics of the fragmentation pattern can be used to give the fatty acyl information of the GPL.

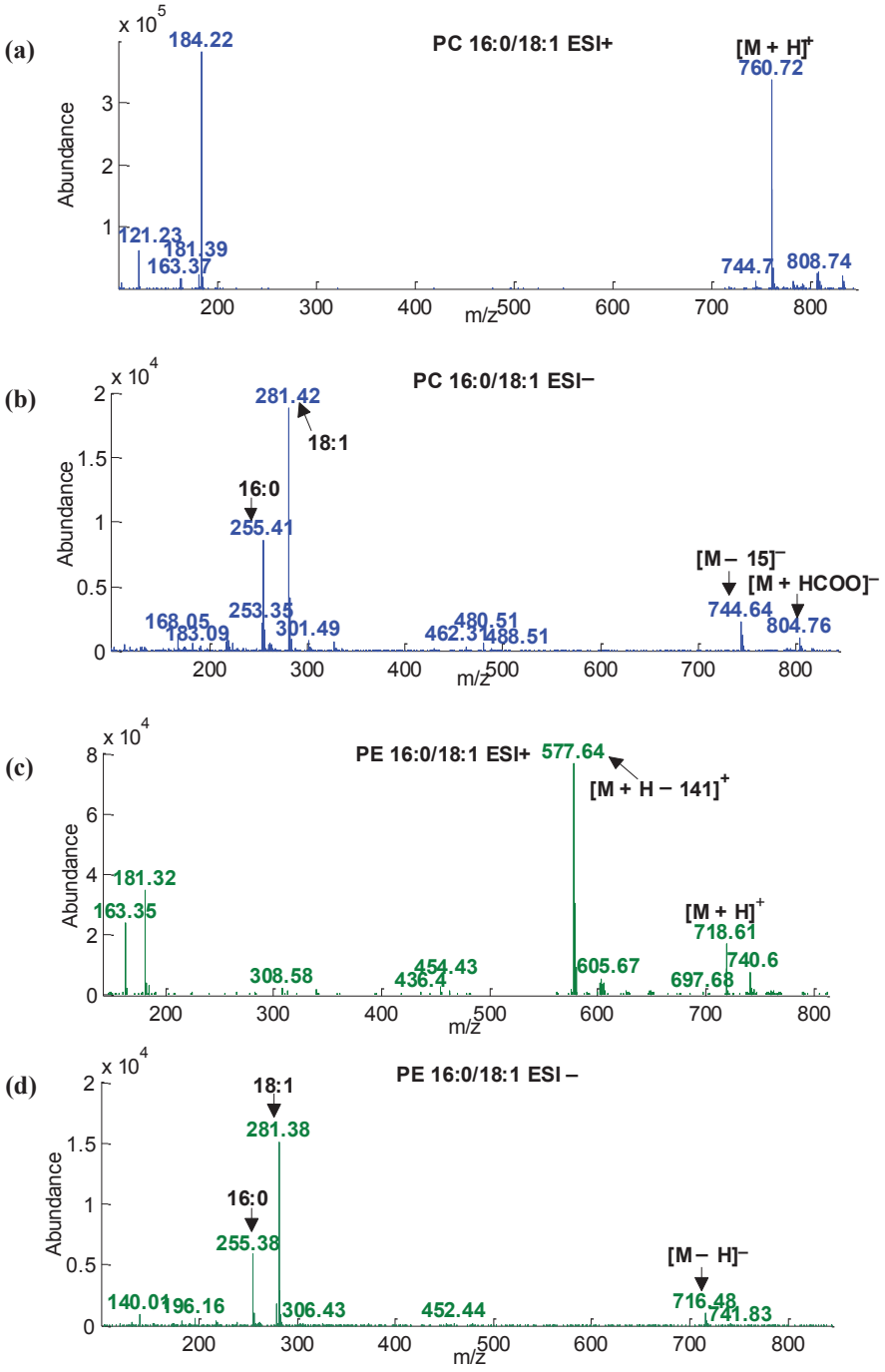


Figure 14 Positive (a) and negative (b) ion ESI mass spectra of PC 16:0/18:1 standard; Positive (c) and negative (d) ion ESI mass spectra of PE 16:0/18:1 standard. Experimental data from QqQ MS.

2.1.4 Triple quadrupole mass spectrometry in phospholipid analysis

The triple quadrupole (QqQ) mass spectrometer has a significant impact on the progress of lipidomics. The basic principle of the QqQ mass spectrometer is as follows: the ions generated from the previous ionization process enter into the first quadrupole where they are filtered according to their *m/z*. The mass separated ions then pass into the collision cell (often a hexapole) where they either undergo CID or pass unhindered to the second quadrupole. The fragment ions are then filtered by the second quadrupole (Figure 15) [84].

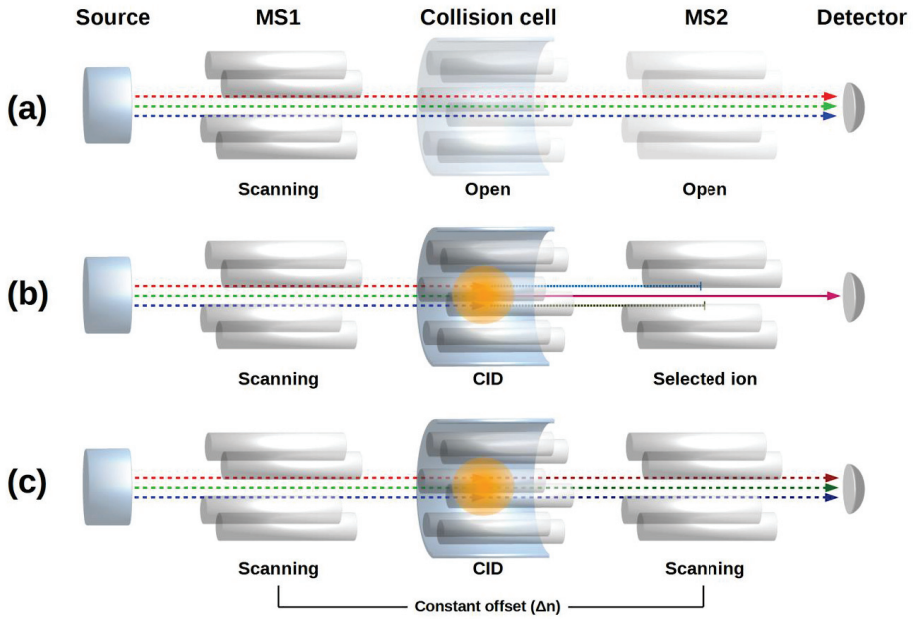


Figure 15 Schematic overview of a QqQ mass spectrometer and different scanning modes. (a) Full MS1 mode, (b) precursor ion scan and (c) neutral loss scan.

Examples of the most common scan techniques employed for lipidomics analyses on a QqQ mass spectrometer are shown in Figure 15. In full MS1 mode (Figure 15a), only one of the quadrupoles is actively used as a mass filter, which is directly analogous to using a single quadrupole mass spectrometer. The mass spectra of PC 16:0/DHA generated by full MS1 scan in negative and positive ion modes are shown in Figure 16a and b, where the full MS1 scan in positive ion mode provides the information of $[M+H]^+$ and $[M+Na]^+$, and the full MS1

scan in negative ion mode provides the corresponding fatty acyl information, *i.e.*, $[R_xCOO]^-$ fragments.

In precursor ion scan (PIS), the first mass analyzer (MS1) operates in scanning mode, separating the molecular ions according to their masses. Then the molecular ions enter the collision cell sequentially where they undergo CID. Generated fragment ions are transmitted into the second mass analyzer (MS2), where they are filtered according to the user-defined dalton (Da) of fragment ions. Only the precursor ions of user-defined Da of fragment ions can be detected (Figure 15b). The resulting spectrum displays all the precursor ions that produce the user-defined fragment ions. The mass spectra of PC 16:0/DHA generated by PIS of 184 Da in positive mode is shown in Figure 16c, where the $[M+H]^+$ precursor ions that generate a fragment ion of 184 Da are shown.

In neutral loss scan (NLS), both mass analyzers (MS1 and MS2) operate in scanning mode, only the precursor ions that lose a neutral fragment of specific mass (Δn) can be detected (Figure 15c). The resulting spectrum shows all the precursor ions that lose a neutral fragment of user defined mass. The mass spectra of PE 18:0/DHA generated by NLS of 141 Da in positive mode is shown in Figure 16d, where the abundances of the $[M+H]^+$ precursor ions that lose a neutral fragment of 141 Da are shown.

Table **4** summarizes some class-specific scanning methods for phospholipids [5, 85, 86]. Generally, the detection of specific subclasses of phospholipids depends on the exploitation of information from head group fragmentation. For example, fragmentation of the choline-containing PC and sphingomyelin species in positive ion mode results in a characteristic phosphocholine headgroup peak (184 Da) [86]. Therefore, PIS of 184 Da will scan for the precursor ions that produce the fragment ion of 184 Da, only detecting PC and sphingomyelin species in this mode. PC and sphingomyelin species are more commonly analyzed in positive ion mode due to high degree of ionization in this mode [87]. Similarly, fragmentation of PE and lyso PE in positive mode normally generates a neutral phosphoethanolamine head group (141 Da) fragment, thus NLS of 141 Da will only detect the PE species that have this characteristic neutral loss [86]. PE can be analyzed both in positive and negative mode. Higher sensitivity can be achieved in positive ion mode with the exception of ether-linked species (plasmanyl and plasmenyl species) [87].

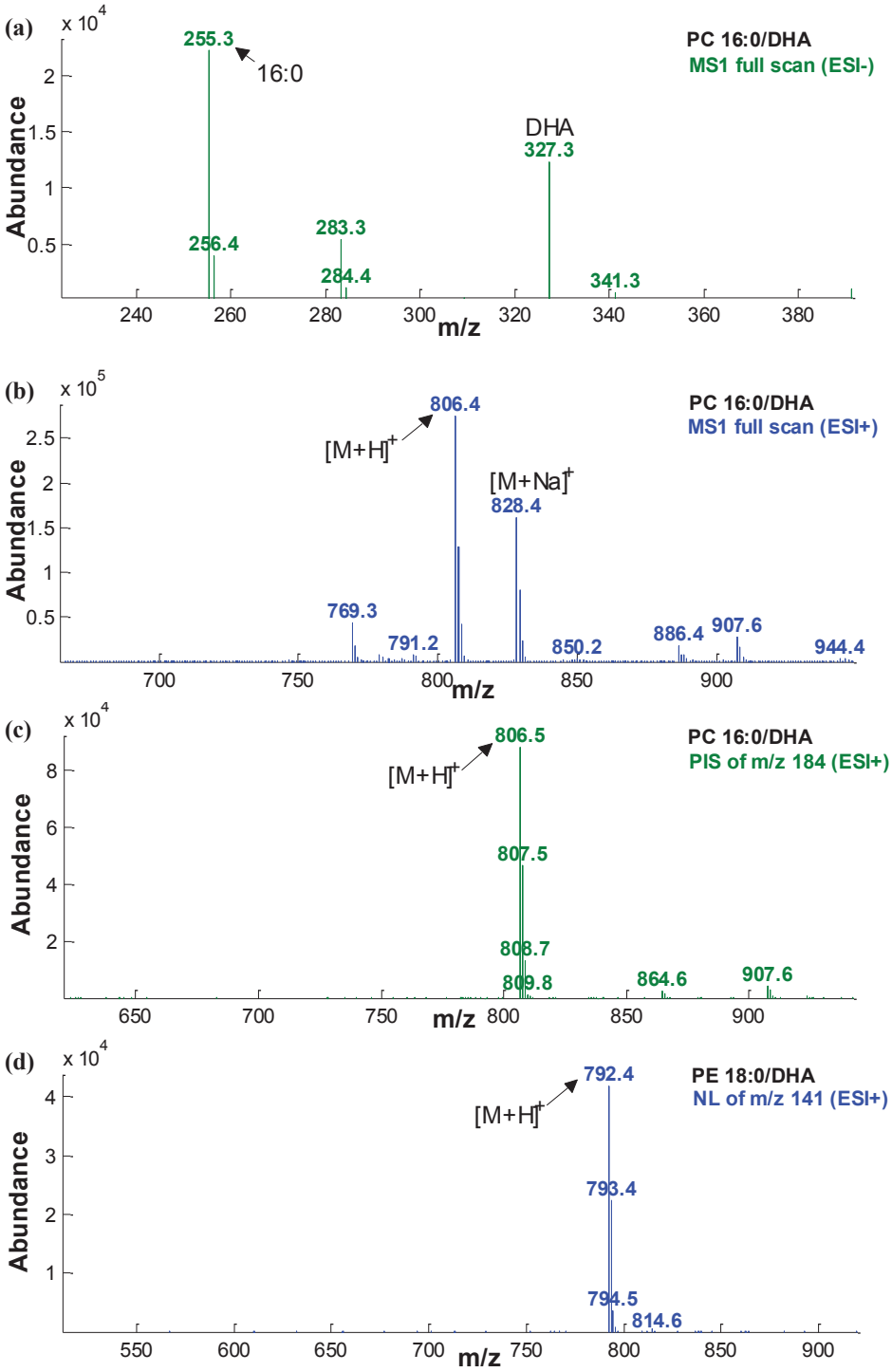


Figure 16 Examples of typical mass spectra from different scan modes. Experimental data from QqQ MS.

Table 4 Examples of typical scanning methods for phospholipid analyses [5, 85, 86]

Subclass	Polarity	Precursor ions	Scanning mode
PC Sphingomyelin	Positive	$[M+H]^+$	PIS of 184 Da
PE	Positive	$[M+H]^+$	NLS of 141 Da
	Negative	$[M-H]^-$	PIS of 196 Da
PS	Negative	$[M-H]^-$	NLS of 87 Da
	Negative	$[M-H]^-$	PIS of 153 Da
	Positive	$[M+H]^+$	PIS of 185 Da
PI	Negative	$[M-H]^-$	NLS of 241 Da PIS of 153 Da
PG	Negative	$[M-H]^-$	PIS of 153 Da PIS of 227 Da
PA	Negative	$[M-H]^-$	PIS of 153 Da
Ceramide	Positive	$[M+H]^+$	PIS of 264 Da

2.1.5 Ion-trap mass spectrometry in triacylglycerol analysis

The ion trap mass spectrometer is another widely employed mass analyzer. It consists of two end-cap electrodes and a ring electrode (Figure 17). The traps have mass selective detection, storage and ejection capabilities. Ions are injected in the trap through one of the end-cap electrodes and ejected through the other end-cap electrode. The ion trap can keep the ions trapped in the electric fields. The orbits of ions are made unstable in a mass-selective manner by increasing the radio frequency voltage that is applied to the device. By adjusting the voltages, ions can be ejected for detection or isolation of ions of a specific *m/z* ratio. The isolated ions can be subjected to CID to produce fragment ions that can be trapped in the analyzer (MS^2). Fragment ions can be ejected for detection or another cycle of ion isolation and subsequent CID process (MS^3). This process can be repeatedly performed and lead to MS^n. This feature of the ion trap analyzer makes it suitable for detailed structural elucidation of compounds.

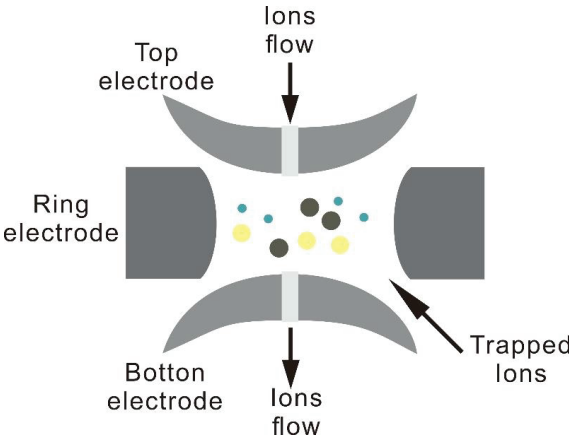


Figure 17 Schematic overview of an ion-trap mass spectrometer [88].

Examples of ESI-MS^2 spectra of ammoniated TAG standards produced by an ion-trap mass spectrometer are shown in Figure 18. A TAG molecule with the same fatty acid on its backbone, such as LnLnLn ('Ln': α-Linolenic acid), exhibits a very simple mass spectrum (Figure 18a) with only a single DAG fragment ion ($[LnLn]^+$ at *m/z* 595.4) resulting from the dissociation of α-linolenic acid from the LnLnLn. A different pattern arises from a TAG molecule containing three different acyl groups such as AOP ('A': arachidic acid, 'O': oleic acid, 'P': palmitic acid). The AOP ammoniated precursor $[M+NH_4]^+$ at *m/z* 907 (Figure 18b) gives rise to three DAG fragments $[PO]^+$, $[AP]^+$ and $[AO]^+$ at *m/z* 577.5, 607.6 and 633.6 respectively. The least abundant DAG fragment ion, at *m/z* 607.6, corresponds to the loss of oleic acid (18:1) from the middle position (*sn*-2), indicating that the cleavage from this particular position is energetically less favoured than the outer positions (*sn*-1 and *sn*-3). Similarly, the mass spectrum of APO (Figure 18c) displays the same three DAG fragment ions observed in the mass spectrum of its stereoisomer AOP. However, the relative intensities of the generated DAG fragments differ between the spectra. In the case of APO (Figure 18c), the DAG fragment $[AO]^+$ at *m/z* 633.6 displays the lowest intensity, indicating the loss of palmitic acid (16:0) from the *sn*-2 position. The observed preferential cleavage of the fatty acids from the outer positions and the relative low intensity at the middle position of the DAG fragments, which enables assigning a particular fatty acid to the *sn*-2 position, have been previously reported [89-91].

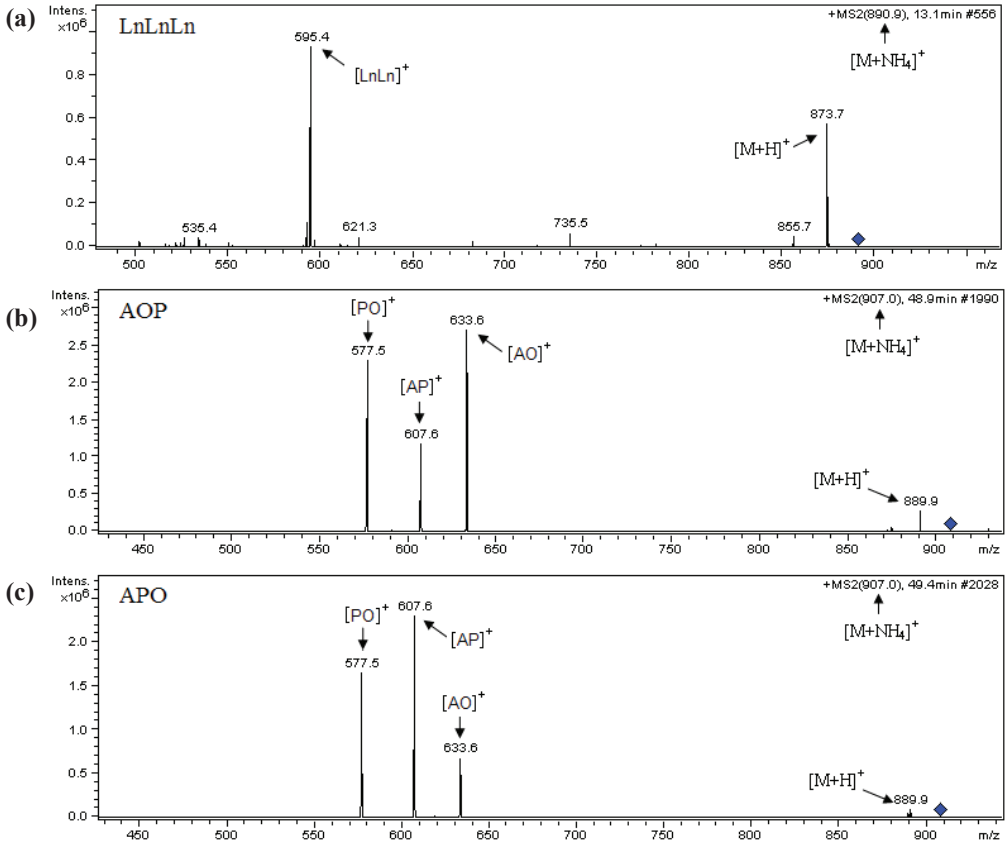


Figure 18 ESI-MS^2 spectra of the ammoniated TAG standards produced by ion-trap mass spectrometer (a) Trilinolenin (LnLnLn), (b) 1-Arachidin-2-Olein-3-Palmitin-glycerol (AOP) and (c) 1-Arachidin-2-Palmitin-3-Olein-glycerol (APO) [81].

2.1.6 Isotope distribution

Most elements occur in nature as a mixture of isotopes. Isotopes are atom species of the same chemical element that have different masses. A list of common elements from lipid molecules and their naturally occurring isotopes is given in Table 5. Lipid molecules are composed of various isotopic elements. Each monoisotopic protonated ion ($[M+H]^+$) used for accurate mass determinations is always accompanied by additional isotope ions. The abundance of the isotope ions ($[M+H+1]^+$, $[M+H+2]^+$, $[M+H+3]^+$) varies with the actual elemental composition. Therefore, different lipid molecules exhibit various isotope distribution patterns. This feature can be used as a powerful tool in identifying the lipids from mass spectral data.

Table 5 Natural isotopic abundances of some common elements in lipid species [92].

Isotope	Mass (Da)	% Abund.	Isotope	Mass (Da)	% Abund.
^{1}H	1.007825	99.985	^{12}C	12.000000	98.90
^{2}H	2.014102	0.015	^{13}C	13.003355	1.10
^{14}N	14.003074	99.63	^{31}P	30.973763	100
^{15}N	15.000109	0.37			
^{16}O	15.994915	99.76	^{32}S	31.972072	95.02
^{17}O	16.999131	0.038	^{33}S	32.971459	0.75
^{18}O	17.999159	0.20	^{34}S	33.967868	4.21

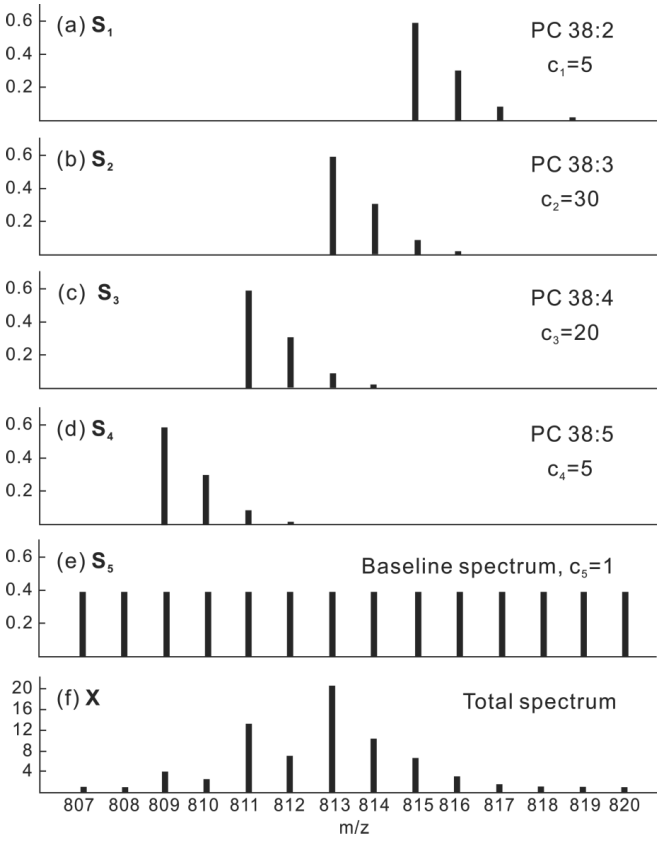


Figure 19 Simulated unit-resolution PIS mass spectra (a) PC 38:2, (b) PC 38:3, (c) PC 38:4, (d) PC 38:5, (e) baseline spectrum and (f) total spectrum generated from (a)–(e). Modified from [75].

In the total mass spectrum of lipid species (either acquired from direct infusion or extracted from LC-MS) obtained from a low mass resolution analyzer, the isotope patterns of different lipids species are often severely overlapped, creating difficulties in the elucidation and quantification of the individual lipid species. An example of simulated unit-resolution PIS mass spectra of several PC species is shown Figure 19. The isotope distribution patterns of these PC species overlap each other. For example, the $[M+H]^+$ signal from PC 38:2 will be influenced by $[M+H+2]^+$ from PC 38:3, and $[M+H+3]^+$ and $[M+H+4]^+$ will also be affected by other signals. Therefore, different isotope correction strategies have been proposed to solve the overlapping issues of isotope patterns [93-95].

2.1.7 Low and high mass resolution mass spectrometry

Mass resolution usually refers to the ability of separating two narrow mass spectral peaks. The IUPAC definition of resolution in mass spectrometry expresses this value as $m/\Delta m$, where m is the mass of the ion of interest and Δm is the peak width (peak width definition) or the spacing between two equal intensity peaks with a valley between them no more than 10 % of their height (10 % valley definition) [96]. The definition and method of measurement of Δm should be reported. A common standard is the definition of resolution based upon Δm being full width of the peak at half its maximum (FWHM) height. Mass resolving power is defined as the measure of the ability of a mass spectrometer to provide a specified value of mass resolution. The procedure by which Δm is defined and measured, and the m/z value at which the measurement was made, should be reported [96].

There is some confusion about the terms of mass resolution and mass resolving power. For example, it has been suggested that resolution should be given by Δm and resolving power by $m/\Delta m$ [97]; however, these definitions are not widely used. The majority of the mass spectrometry community uses resolution as defined by IUPAC in 2013 [96]. The term resolving power is not widely used as a synonym for resolution.

Low-resolution mass analyzers, such as ion trap and QqQ, can reach the resolving power ($m/\Delta m_{50\%}$) of 3000 [98]. High-resolution mass analyzers are specifically those capable of routine broadband resolving power ($m/\Delta m_{50\%}$) >10,000 and of high mass accuracy (<5 ppm), such as linear trap quadrupole (LTQ) Orbitrap, quadrupole time-of-flight (Q-TOF) and Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometers. For example, LTQ-Orbitrap has resolving power ($m/\Delta m_{50\%}$) up to 150,000 or higher [99], and FT-ICR has resolving power ($m/\Delta m_{50\%}$) up to 1,000,000 or higher [100].

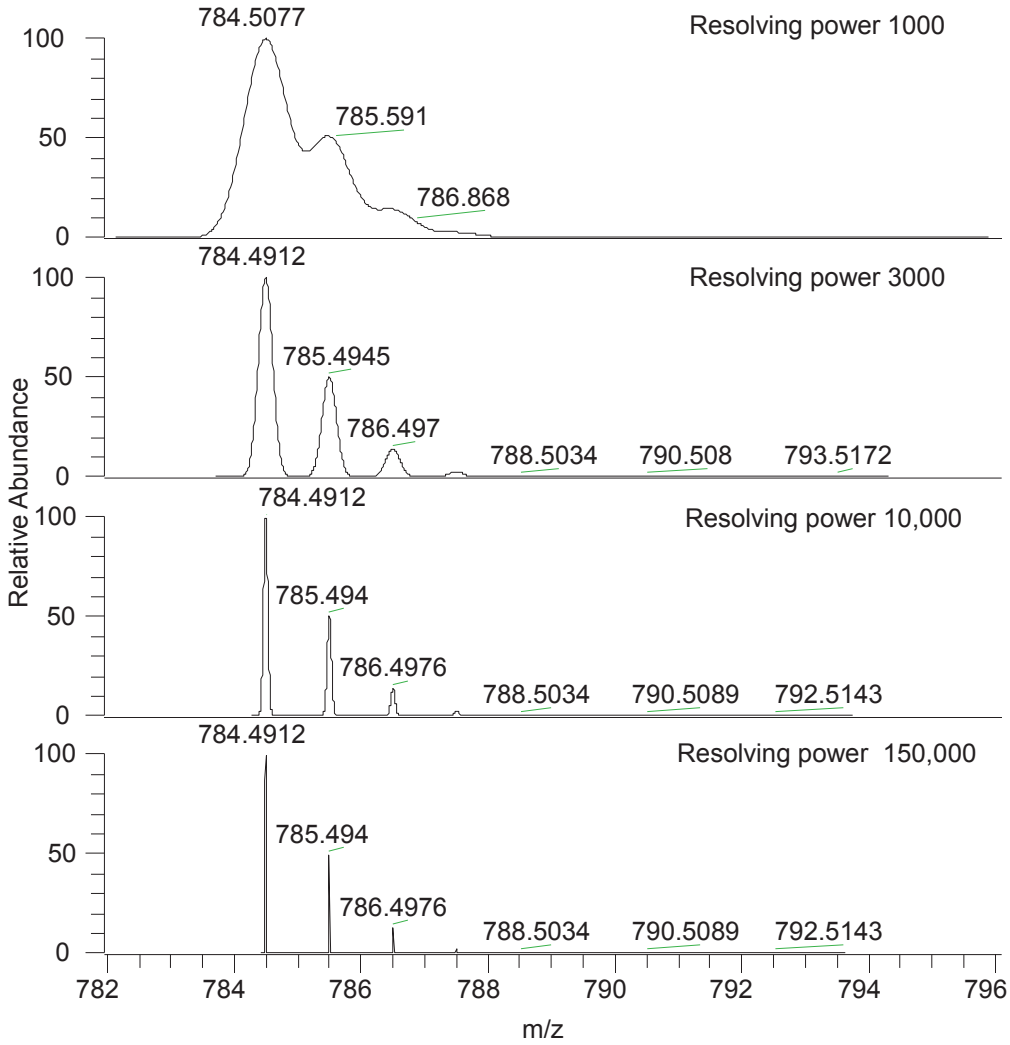


Figure 20 The isotope pattern of predicted continuous spectra of positive ion $[M+H]^+$ of PE 18:4/22:6 ($C_{45}H_{70}NO_8P$) using different levels of mass resolving power ($m/\Delta m_{50\%}$) from 1000 to 150,000 generated by the Xcalibur software (Thermo Fisher Scientific).

As discussed in the previous section, every isotope of every element has a different mass defect (*i.e.*, difference between exact mass and the nearest integer). For example, the mass difference between the lightest stable isotope of carbon (^{12}C) and its next heavy counterpart (^{13}C) is 1.00335 Da, and the difference between the lightest stable nitrogen (^{14}N) and its successor (^{15}N) is 0.99703 Da [92]. These fine differences between the isotopes of different elements can be distinguished by modern high-resolution mass analyzers with sufficient

accurate mass measurement, and thus exact elemental composition (*e.g.* $C_cH_hN_nO_oP_p$) can be determined. However, if the mass differences between isomers are too small (*e.g.* $\sim 10^{-9}$ Da), it will still be problematic to measure in the mass spectrum [97]. In addition, there are a lot of structural lipid isomers (*e.g.* skeletal, positional and functional isomers), for which high-resolution mass measurements cannot yield a unique composition. Thus MS/MS or LC-MS is needed to provide additional information.

Figure 20 shows an example of predicted continuous spectra of positive ion $[M+H]^+$ of PE 18:4/22:6 ($C_{45}H_{70}NO_8P$) using different levels of mass resolving power ($m/\Delta m_{50\%}$) from 1000 to 150,000. It can be seen that the isotope pattern of PE 18:4/22:6 is completely overlapped when the mass resolving power ($m/\Delta m_{50\%}$) is 1000. As the resolving power ($m/\Delta m_{50\%}$) increases to 3000, 10,000, and further to 150,000, the isotope pattern becomes well distinguished and the mass spectral peaks become sharper.

2.2 Chemometric techniques

2.2.1 Component detection algorithm (CODA)

LC-MS employing ESI or APCI ionization techniques is a powerful analytical tool for the specific detection and potential identification of compounds in complex mixtures. However, analysts often encounter several problems when analyzing LC-MS data. First, the amount of data obtained from LC-MS has increased a lot over the years as a result of instrumental developments and applications dealing with increasingly complex multi-component samples and matrices. Interpretation of large data sets has thus become a formidable challenge. Secondly, the combination of LC with MS, particularly using ESI as an ionization method, can result in chromatograms and mass spectra with a high level of background and noise, which comes from a number of sources, such as the LC mobile phase and buffers [101]. The contributions of solvent and background will often dominate the chromatogram. As a consequence of this, the resulting total ion chromatograms do not exhibit a clear separation of the sample mixture. Even though great improvements have been made in the instrumentation, the output signals still contain a large portion of noise embedding the useful information together with false peaks and spikes [102]. Due to the above mentioned problems, the data interpretation can be much more time-consuming than the data acquisition and has become a bottleneck in LC-MS method development and analysis.

Currently, several data pretreatment techniques are available to assist the data analysis generated from LC-MS. These are mainly focused on the preprocessing of data, such as reducing the noise and background, improving the signal to noise ratio (S/N), extracting high quality chromatograms, *etc*. Singular value decomposition (SVD) is a commonly used data reduction algorithm to improve the S/N in LC-MS data [103]. Sequential paired covariance (SPC) [104], higher order-sequential paired covariance (HO-SPC) [105], windowed mass selection method (WMSM) [106], and component detection algorithm (CODA) [101, 106-109] are chemometric methods developed specifically for chromatography/MS data [106]. CODA was specifically developed for LC-MS data in order to reduce random noise and high background by selecting only high-quality (low noise and background) mass chromatograms from complex LC-MS data [101, 107]. The method has been demonstrated to be effective in the analysis of LC-MS data from urine samples, human serum and peptides [106, 109, 110]. CODA is based on two factors: 1) background mass chromatograms will have a high mean value compared to good mass chromatograms; 2) noisy, spiky mass chromatograms will be more affected by smoothing than good mass chromatograms. Consequently, by calculating a similarity coefficient (so-called mass chromatographic quality, MCQ) between the original and smoothed chromatogram for each mass, the algorithm will distinguish spiky chromatograms and solvent chromatograms from the mass chromatograms of potential components. The noisy mass chromatograms will be easily eliminated by selecting masses with high similarity coefficients, thus significantly increasing the quality in the LC-MS data.

Briefly, the process of CODA consists of the calculation of the similarity index by the following equation:

$$c_j = \frac{1}{\sqrt{r-w}} \sum_{i=1}^{r-w+1} \mathbf{X}_{ij}\mathbf{X}^*_{ij} \qquad \text{Equation 2}$$

Where c_j is the similarity index (also referred as MCQ) for j^{th} mass chromatogram, *i.e.*, mass chromatographic quality value, r is the number of rows (scans) in the original dataset, w is the width of a rectangular smoothing window**,** $\mathbf{X}_{ij}$ stands for an element of data matrix at time point i and mass channel j, $\mathbf{X}^*_{ij}$ represents the smoothed and standardized data matrix.

The MCQ in Equation 2 is calculated for each mass chromatogram. A high similarity index indicates that the mass chromatogram in question contains information about eluting components, while a low similarity index shows that any signal at this mass may be caused by

noise or background. By specifying a MCQ threshold (between 0 and 1), the high-quality mass chromatograms will be retained by selecting the mass chromatograms with MCQ values higher than the defined threshold, while other mass chromatograms with lower MCQ values will be discarded.WMSM and CODA are optimal tools for LC-MS data analysis. The former is more suitable for mass chromatograms with frequent and high intensity noise peaks, and the latter performs well for data with high background signals [106].

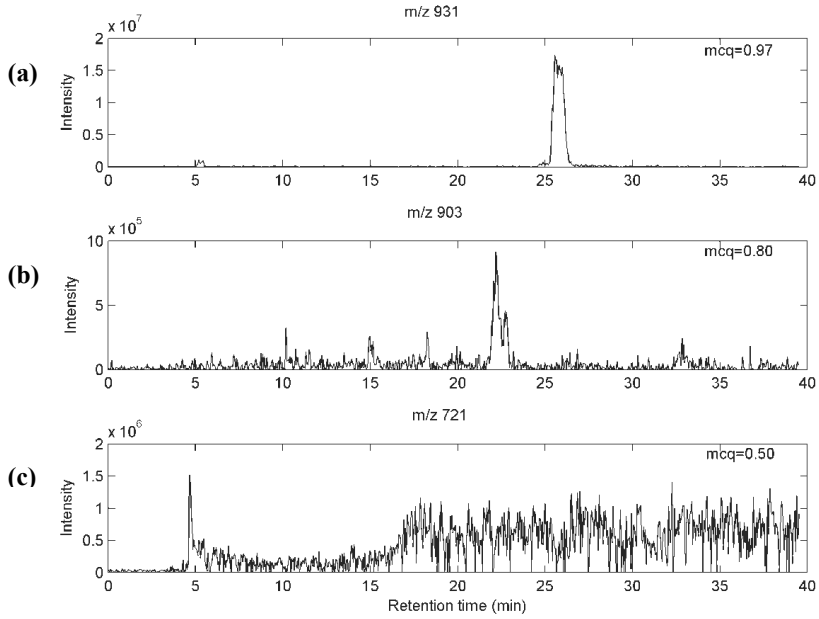


Figure 21 Typical mass chromatograms of TAG standards mixture from LC-ESI-MS experiments in decreasing order of MCQ. Experimental data from Ion-trap MS.

The performance of CODA was tested and verfied on the data from a TAG standards mixture. Examples of three typical mass chromatograms with different MCQ values from LC-ESI-MS data of the TAG mixture are shown in Figure 21. For the mass chromatogram with noise and background, the process of data smoothing and standardization will make the values of $\mathbf{X}_{ij}$ and $\mathbf{X}^*_{ij}$ significantly different, leading to a low MCQ value (Figure 21c), while the high-quality mass chromatogram remains nearly unaffected during the process and gives rise to a similar $\mathbf{X}^*_{ij}$ value as the original $\mathbf{X}_{ij}$ value, resulting a MCQ value close to 1 (Figure 21a). Therefore, only mass chromatograms with MCQ values higher than an assigned threshold will be selected, while others chromatograms are discarded. The typical MCQ value used is 0.89

[101, 107]. However, the MCQ threshold needs to remain variable when processing different data. For example, when the concentrations of the components in a sample are very low, the mass chromatograms will have a lower quality. In this case a lower MCQ threshold could be defined in order to detect all components present. Therefore, the assignment of the MCQ threshold needs to be examined individually according to the features of each dataset.

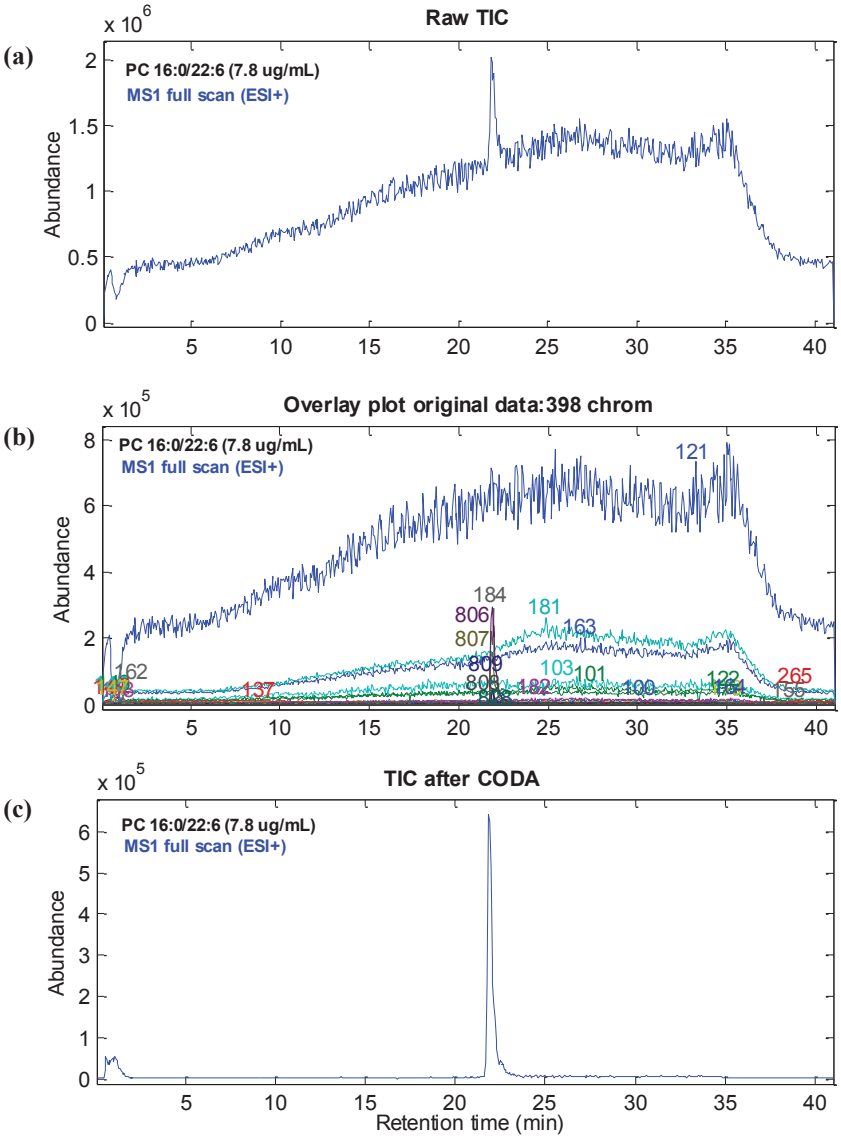


Figure 22 LC-ESI-MS analysis of PC 16:0/22:6 (concentration 7.8 µg/mL). (a) The original TIC profile; (b) The overlay plot of all 398 mass chromatograms; (c) The TIC profile resulting from 6 mass chromatograms selected by CODA (window size = 5, MCQ = 0.89). Experimental data from QqQ MS.

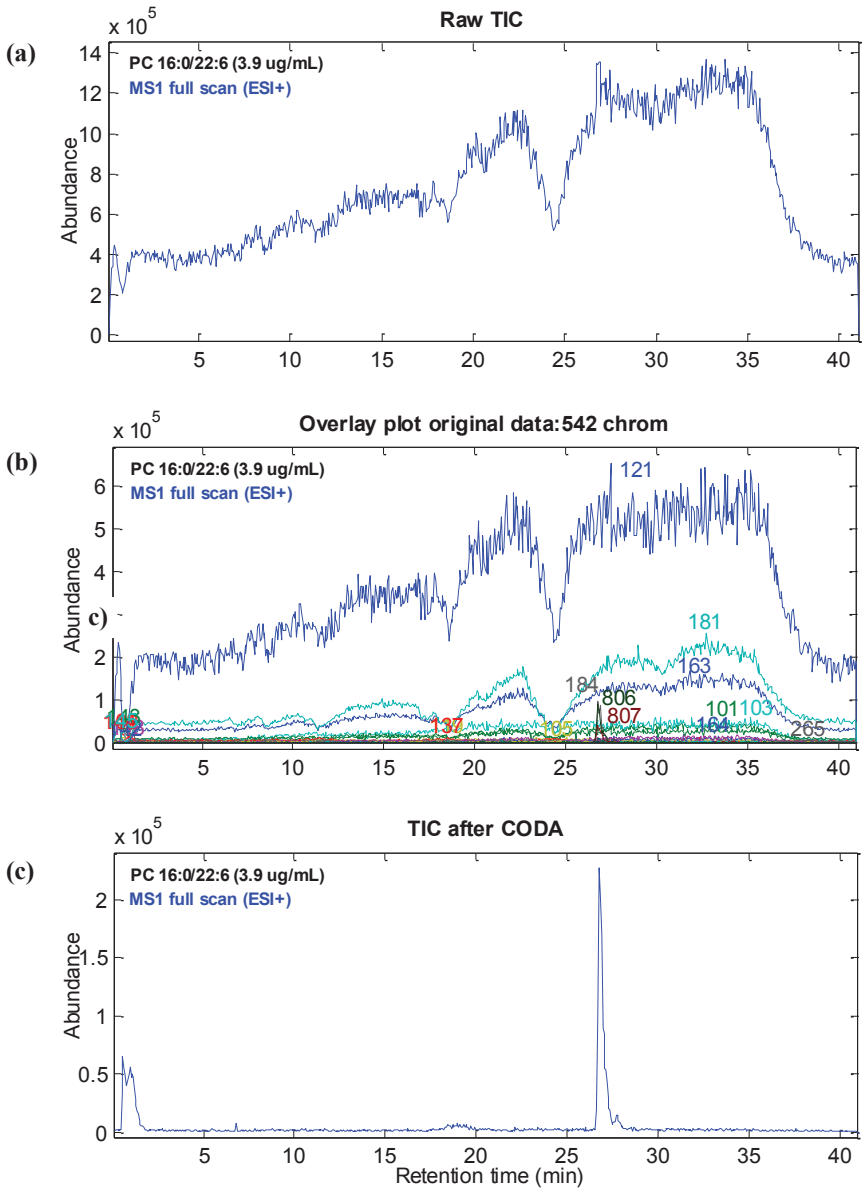


Figure 23 LC-ESI-MS analysis of PC 16:0/22:6 standard (concentration 3.9 µg/mL). (a) The original TIC profile; (b) The overlay plot of all 398 mass chromatograms; (c) The TIC profile resulting from 8 mass chromatograms selected by CODA (window size = 5, MCQ = 0.89). Experimental data from QqQ MS.

By applying CODA, the useful information will be extracted effectively from the complex LC-MS data sets, greatly reducing the noise and background and eliminating their interferences. The following examples will show how CODA extracts the useful information from LC-MS data. The original TIC and the overlay plot of all 398 mass chromatograms derived from a PC 16:0/22:6 standard analysis by LC-MS are shown in Figure 22a and b, respectively. The original TIC is quite nosiy (Figure 22a), which is ascribed to a high level of background from the low-quality chromatograms, *e.g.*, *m/z* 121, 181, 163 (Figure 22b). After the application of CODA, only 6 high-quality mass chromatograms, *e.g.*, *m/z* 806, 184, 807 (MCQ $\geq$ 0.89), are selected from a total of 398 chromatograms. The noise and high background are greatly reduced as can be observed in the TIC after CODA (Figure 22c).

When the concentration of PC 16:0/22:6 is diluted to 3.9 μg/mL, the peak of PC 16:0/22:6 is not visible in the raw TIC profile due to the high background in the chromatogram (Figure 23a). The overlay plot of all 542 mass chromatograms of the original dataset (Figure 23b) shows that the peak of PC 16:0/22:6 is completely embedded in the background mass chromatograms, such as *m/z* 121, 181, 163. The application of CODA efficiently removes low quality background mass chromatograms and extracts high quality mass chromatograms of the analytes. The peak of PC 16:0/22:6 can thus be clearly visulized in Figure 23c.

2.2.2 Principal component analysis (PCA)

Multivariate analysis refers to a set of statistical techniques used to analyze data that arises from more than one variable. Commonly employed multivariate techniques include principal component analysis (PCA), partial least squares (PLS) and soft independent modeling of class analogy (SIMCA). In the present thesis, emphasis is placed on PCA for the purpose of discrimination and classification.

PCA is a well-documented multivariate method for extracting the most important information from large datasets by reducing the dimensionality of the data. It is very useful when dealing with data from LC-MS and direct infusion MS with a large number of variables. New variables can be constructed by rotating and constructing orthogonal linear combinations of the original variables. These new variables are called “principal components” and ideally account for the majority of the variability in the original data. In this way, the dataset can be represented by relatively few principal components instead of a large number of original variables. The first principal component is the major axis of the points in the *p*-dimensional

space that accounts for maximum amount of variance in the data. The second principal component is perpendicular to the first principal component and explains the largest amount of variation that is not explained by the first principal component. Further principal components can be extracted by repeating the same principle [111]. Once obtained, the scatter plots of the principal components can be used to assess similarities and differences between samples and variables and to determine samples clustering.

The core formula of PCA is displayed below. The data matrix **X** can be decomposed into scores $\mathbf{t}_i$ and loadings $\mathbf{p}_i$. The score and loading vectors are collected in a score matrix **T** and a loading matrix **P** if more than component is extracted.

$$\mathbf{X} = \mathbf{T}\mathbf{P}^{\mathrm{T}} + \mathbf{E} = \mathbf{t}_1\mathbf{p}_1^{\mathrm{T}} + \mathbf{t}_2\mathbf{p}_2^{\mathrm{T}} + \cdots + \mathbf{t}_l\mathbf{p}_l^{\mathrm{T}} + \mathbf{E} \qquad \text{Equation 3}$$

In Equation 3, **X** is an $m \times n$ matrix, consisting of m samples and n variables. The residual matrix **E** has the same dimension. The score matrix **T** has dimensions $m \times l$, and the loading dimensions $n \times l$, where l is the number of calculated principal components, and $^{\mathrm{T}}$ denotes transposition.

The score plot can be used to evaluate cluster patterns, trends and outliers. The loading plot can be used to interpret the relationship among variables, and measure the contribution of the variables to the principal components.

2.2.3 Least squares spectral resolution approach (LSSR)

Progress in the lipidomics field has been driven by advances in instrumental techniques, but the developments in lipidomics is still lagging due to the complexity of lipids and the lack of powerful tools for lipid analyses [112]. Lipidomic platforms are mainly based on MS approaches including shotgun lipidomics and LC-MS [113]. The large amounts of data produced demand powerful computational tools capable of rapidly analyzing lipidomic data. To address this issue, the least squares spectral resolution (LSSR) approach was developed for MS and LC-MS lipidomic data processing, including detection, identification, quantification and visualization of results. This framework is based on least squares resolution of spectra and chromatograms from theoretically calculated isotope distributions of mass spectra, which eliminates the need for isotope correction. The original LSSR approach was developed for the analysis of PC and PE species from low-resolution MS or LC-MS data [75]. Currently, this methodology has been extended in two aspects: i) to support the analysis of high resolution

MS or LC-MS lipidomic data; ii) to cover 38 different lipid classes, including the most common GPL (PC, PE, PA, PI, PS *etc.*), sphingolipids (ceramides, sphingomyelins *etc.*) and glycerolipids (TAG, DAG).

The principle of this methodology is briefly described below and a schematic workflow of the LSSR methodology for the analysis of lipidomic data is presented in Figure 24. The details can be found in Papers **III** and **IV**.

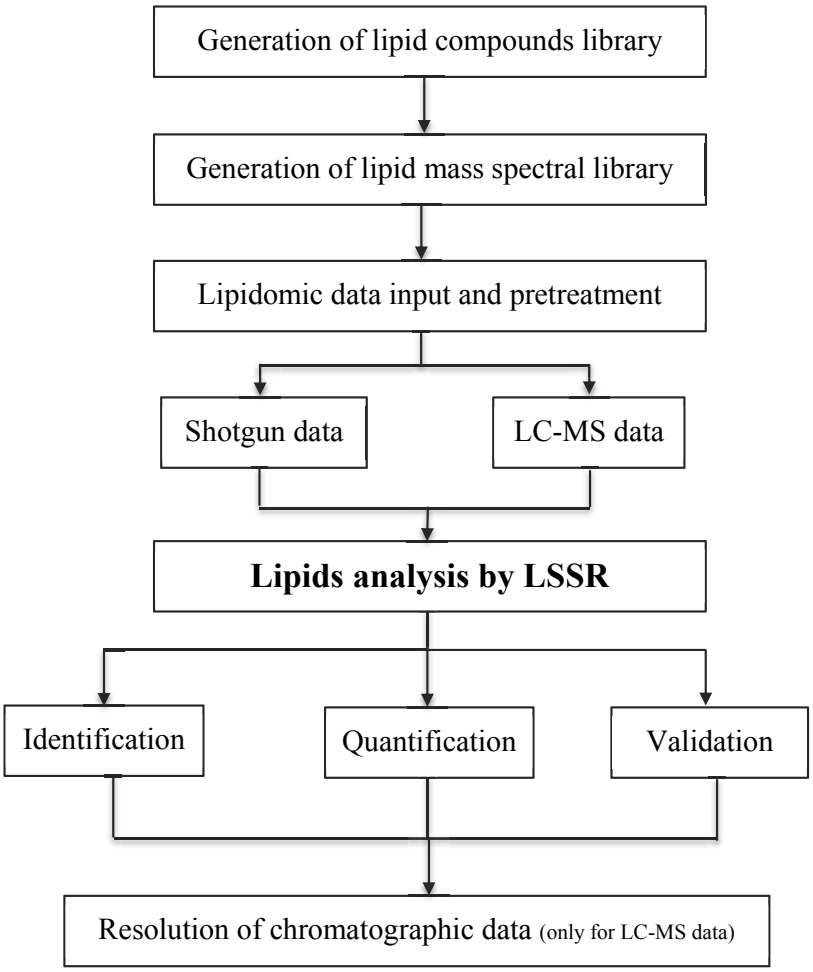


Figure 24 Schematic workflow of the LSSR methodology for the analysis of lipidomic data.

1) Generation of lipid compounds library

The targeted lipid compound library can be generated based on the structural composition features of various lipid classes, specifically, the list of fatty acids, sphingoid bases, lipid class core formula (head groups). For example, a targeted library of PC species can be generated when the data only contains PC classes.

2) Generation of lipid mass spectral library

Currently supported ion types are $[M+H]^+$, $[M+NH_4]^+$, $[M+Na]^+$, $[M-H]^-$, $[M-H_2O]^-$, $[M+CHO_2]^-$, $[M+C_2H_3O_2]^-$, and $[M+Cl]^-$. By choosing the corresponding ion types, the theoretical mass spectra of targeted lipid classes are originally generated with a mass resolution of 0.001 Da and then binned to the required resolution, which should match the imported experimental data.

3) Lipidomic data input and pretreatment

The program supports the following MS data formats: Network Common Data Form (NetCDF), mzXML, mzData, Agilent Chemstation raw data, and JEOL jmc profiles. These data can be directly imported into the software. A sum spectrum, described by a row vector $\mathbf{x}^T$ is calculated from the entire time range or from a sub-section of the LC-MS chromatogram in the imported data. The data can be pretreated by CODA or other filtering methods.

4) Lipid analysis by LSSR

The sum spectrum (described by a row vector $\mathbf{x}^T$) calculated from Step 3 is compared with the generated mass spectral library, and compounds in the library that do not have a base peak that is found in $\mathbf{x}^T$ above a predefined threshold level are excluded from the library. This gives a condensed library of spectra described by the matrix $\mathbf{S}^T$, where each spectrum is described by a row in $\mathbf{S}^T$. Thereafter, $\mathbf{x}^T$ and $\mathbf{S}^T$ are condensed to contain only masses that are present in both.

A column vector of concentrations, $\mathbf{c}$, of each compound in $\mathbf{S}^T$ is calculated by least squares regression according to Equation 4.

$$\mathbf{c} = \mathbf{x}^T\mathbf{S}(\mathbf{S}^T\mathbf{S})^{-1} \qquad \text{Equation 4}$$

A validation spectrum, $\mathbf{x}^T_{val}$, is thereafter calculated from $\mathbf{c}$ and the spectra in $\mathbf{S}^T$ by Equation 5.

$$\mathbf{x}^T_{val} = \mathbf{c} \times \mathbf{S}^T \qquad \text{Equation 5}$$

If the resolution is accurate, $\mathbf{x}^T$ and $\mathbf{x}^T_{val}$ should be similar. Large deviations may occur if ions in $\mathbf{x}^T$ are influenced by noise or compounds that are not in $\mathbf{S}^T$, or if the abundances of isotopes in $\mathbf{x}^T$ do not match the abundances in $\mathbf{S}^T$, *e.g.* due to non-linearites in the experimental data.

5) Resolution of chromatographic data

Chromatographic data can be expressed by a matrix $\mathbf{X}^T$, where each row represents a spectrum. The concentration profiles of each mass, $\mathbf{C}$, can then be estimated from $\mathbf{S}^T$ by Equation 6.

$$\mathbf{C} = \mathbf{X}^T\mathbf{S}(\mathbf{S}^T\mathbf{S})^{-1} \qquad \text{Equation 6}$$

In this case $\mathbf{X}^T$ must be filtered to contain the same masses as $\mathbf{S}^T$.

The presented methodology has been applied to analysis of lipids extracts from cod, mouse and porcine brain. The flexibility of the methodology allows it to be expanded to support more lipids classes and more data interpretation functions, making it a promising tool in lipidomic data analysis.

3 Triacylglycerols study

This chapter is a summary of results from Papers **I** and **II**. Marine oils are important sources of *n*-3 PUFA that have attracted extensive interests due to their numerous potential health benefits [36-39, 114-118]. Most marine oils consist almost exclusively of TAG esterified with various fatty acids. Structural characterization of TAG species in marine oils represents a great challenge for analysts due to the wide variety of fatty acids and the complexity of naturally occurring TAG species of marine origin. In addition, it is of vital importance to explore the capabilities of various techniques for differentiating marine oils from commercial, nutritional and biochemical points of view. These aspects have been much neglected compared to the study of plant oils. Paper **I** studied the characterization of the naturally occurring TAG in cod liver oil by ion-trap LC-ESI-MS^2. Paper **II** evaluated various fingerprinting strategies for differentiating marine oils based on their TAG composition and four different TAG profiles using liquid chromatography electrospray single and tandem mass spectrometry (LC-ESI-MS and LC-ESI-MS^2) by PCA.

3.1 Characterization of the TAG in cod liver oil by LC-ESI-MS^2

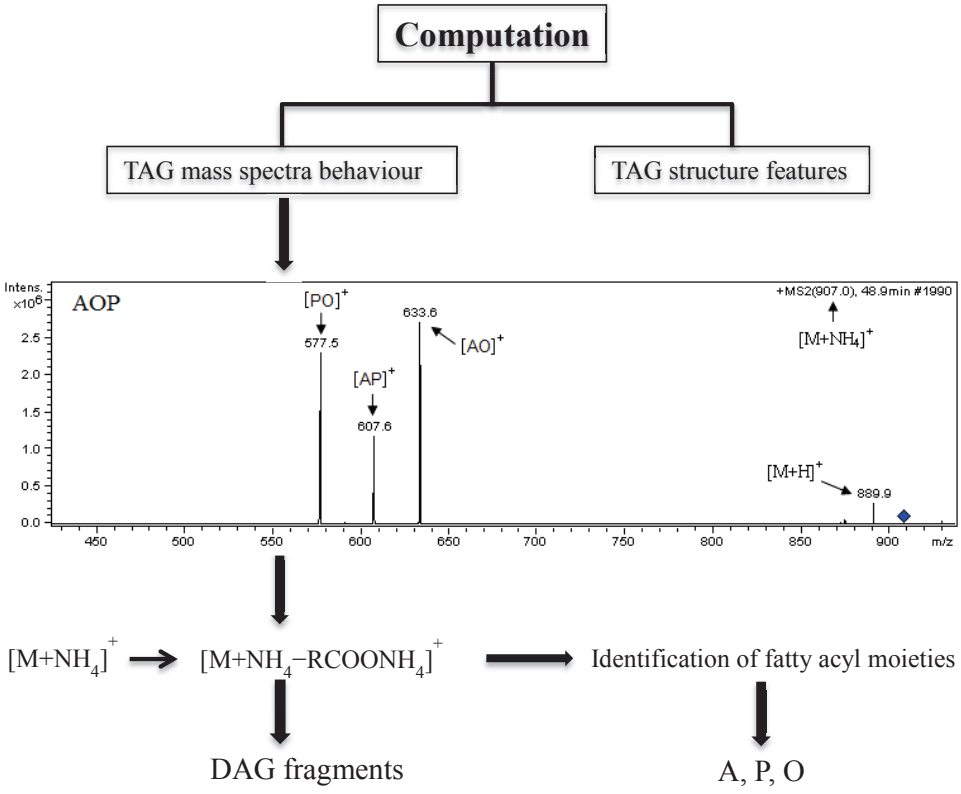
In Paper **I**, an LC-ESI-MS^2 strategy for characterizing intact TAG species in cod liver oil was established, and the corresponding automated interpretation algorithm was developed to assist the automatic interpretation of the relative arrangement of the acyl groups on the glycerol backbone of cod liver oil. The computation theory for the elucidation of a TAG molecule is based on the behaviour of the mass spectra of TAG and the basic structural features of a TAG molecule, as illustrated in Figure 25 and Figure 26. The behaviour of ESI-MS^2 mass spectra of TAG was demonstrated in previous studies [89-91] and verified by using TAG standards in our study. Briefly, the precursor adduct ion from the ESI-MS^2 mass spectrum of TAG produces very abundant DAG fragment ions due to the loss of fatty acyl moieties from the glycerol backbone. The fatty acid that corresponds to the least abundant DAG fragment (lowest intensity) will be assigned to the *sn*-2 position on the TAG backbone.

A mass spectrum of an ammoniated TAG (AOP) is shown in Figure 25 (‘A’: arachidic acid, ‘O’: oleic acid, ‘P’: palmitic acid), where different DAG fragments are generated due to the loss of fatty acyl moieties from the glycerol backbone.

Thus, all the fatty acyl moieties in the mass spectrum can be calculated by Equation 7.

$$FA_i = [M+NH_4]^+ - Frag_i - 17 \qquad \text{Equation 7}$$

Where $[M+NH_4]^+$ represents the masses of ammoniated TAG adduct, FA_i denotes the masses of fatty acyl moieties, and $Frag_i$ denotes the masses of DAG fragments. The identified fatty acyl moieties can be combined on the glycerol backbone of TAG, which yields 12 possible TAG molecules in this case (Figure 25). The theoretical total number of CH_2-CH_2 groups (*a*) and the total number of CH_2=CH_2 groups (*b*) can be easily obtained.



AOP	APO	PAO	AAA	PPP	OOO
AAP	APP	AOO	AAO	OOP	OPP

Figure 25 Computation theory based on the TAG mass spectra behaviour and the basic structural features of a TAG molecule. Example of elucidating an ammoniated TAG (AOP) molecule is shown.

The basic structural features of a TAG molecule are shown in Figure 26. Different TAG molecules possess several common chemical groups. These common structural features of a TAG molecule can be used to generate a general algebraic expression for TAG elucidation as shown in Equation 8.

$$[M+NH_4]^+ = 236 + 28a + 26b \qquad \text{Equation 8}$$

Where a represents the total number of CH_2-CH_2 groups and b represents the total number of CH_2=CH_2 groups. This equation can be converted to Equation 9 shown below.

$$a = \frac{[M+NH_4]^+ - 236 - 26b}{28} \qquad \text{Equation 9}$$

By introducing the experimental $[M+NH_4]^+$ value and substituting automatically only integral numbers of b from 0 to 18 (the total possible range of double ethenyl bonds), it is possible to estimate a, the total number of single ethylene bonds, by using Equation 9. A positive TAG identification can be achieved when the theoretical a and b values of the possible TAG molecules are equal to values estimated from Equation 9.

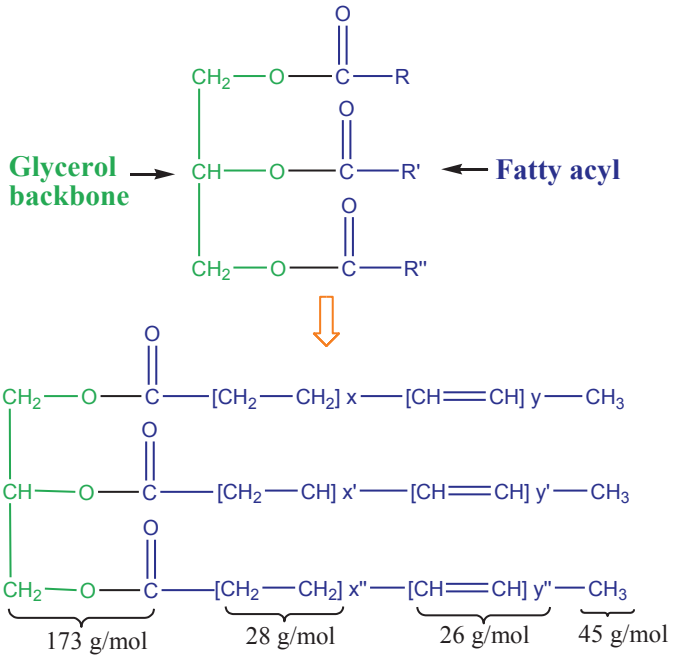


Figure 26 The basic structural features of a TAG molecule. A TAG molecule can be decomposed into several chemical groups. A general expression for TAG elucidation can be generated as shown in Equation 8.

The methodology was tested and verified by using TAG standards. The relative arrangement of the acyl groups on the intact molecules in cod liver oil was identified and the elucidation of TAG positional and structural isomers in cod liver oil was demonstrated. The fully identified TAG species with their *sn*-2 and *sn*-1/3 positions are presented in Paper **I** (no distinction is made between the *sn*-1/3 positions). The results showed that the fatty acids exhibiting the highest relative concentrations by the lipase method [81], *i.e.*, 16:0, 16:1, 18:1, 20:1, 22:1, EPA and DHA, were most frequently detected in the various TAG structures by LC-ESI-MS^2.

3.2 Fingerprinting strategies for differentiating marine oils based on TAG profiles

Paper **II** evaluated the capabilities of fingerprinting several marine oils (seal, whale, cod liver and salmon) and plant oils (soy, rapeseed and linseed) based on their TAG composition and four different types of TAG profiles from LC-ESI-MS and LC-ESI-MS^2 experiments by means of PCA. Four types of TAG profiles, namely, total ion current (TIC) and total mass spectral (TMS) profiles from LC-ESI-MS and LC-ESI-MS^2 experiments were used to investigate the performance of different TAG profiles in the marine oils discrimination and to find the most appropriate profiles that represent the characteristics of TAG patterns [119].

The general workflow is illustrated in Figure 27. The LC-MS dataset is a 2-D matrix $\mathbf{X}_{m\times n}$ where each row represents an object (a mass spectrum at a scanning point) and each column represents a variable (a chromatogram at a *m/z* ratio). $\mathbf{X}_{m\times n}$ can be represented by the following equation.

$$\mathbf{X}_{m\times n} = \mathbf{C}_{m\times p}\mathbf{S}^{\mathbf{T}}_{n\times p} + \mathbf{E} \qquad \text{Equation 10}$$

where m is the number of scanning points, n is the number of *m/z* ratios, p designates the number of components in the system, $\mathbf{C}_{m\times p}$ and $\mathbf{S}_{n\times p}$ stand for the chromatographic concentration profiles and mass spectral profiles, respectively. The superscript **T** represents the matrix or vector transposition and **E** is the matrix of noise. The TIC profiles are obtained by collecting the contributions of all components in the $\mathbf{X}_{m\times n}$ from the chromatographic direction, *i.e.*, by plotting the summed current from each mass spectrum against the scanning points (retention time). In this case, only chromatographic information is retained. The TMS profiles, on the other hand, collect all contributions of components in the dataset $\mathbf{X}_{m\times n}$ from the mass spectral direction, *i.e.*, by plotting the sum of all the ion intensities against the *m/z*

values. After eliminating the noise and background of the four profiles by using CODA, the pre-processed profiles are evaluated by PCA.

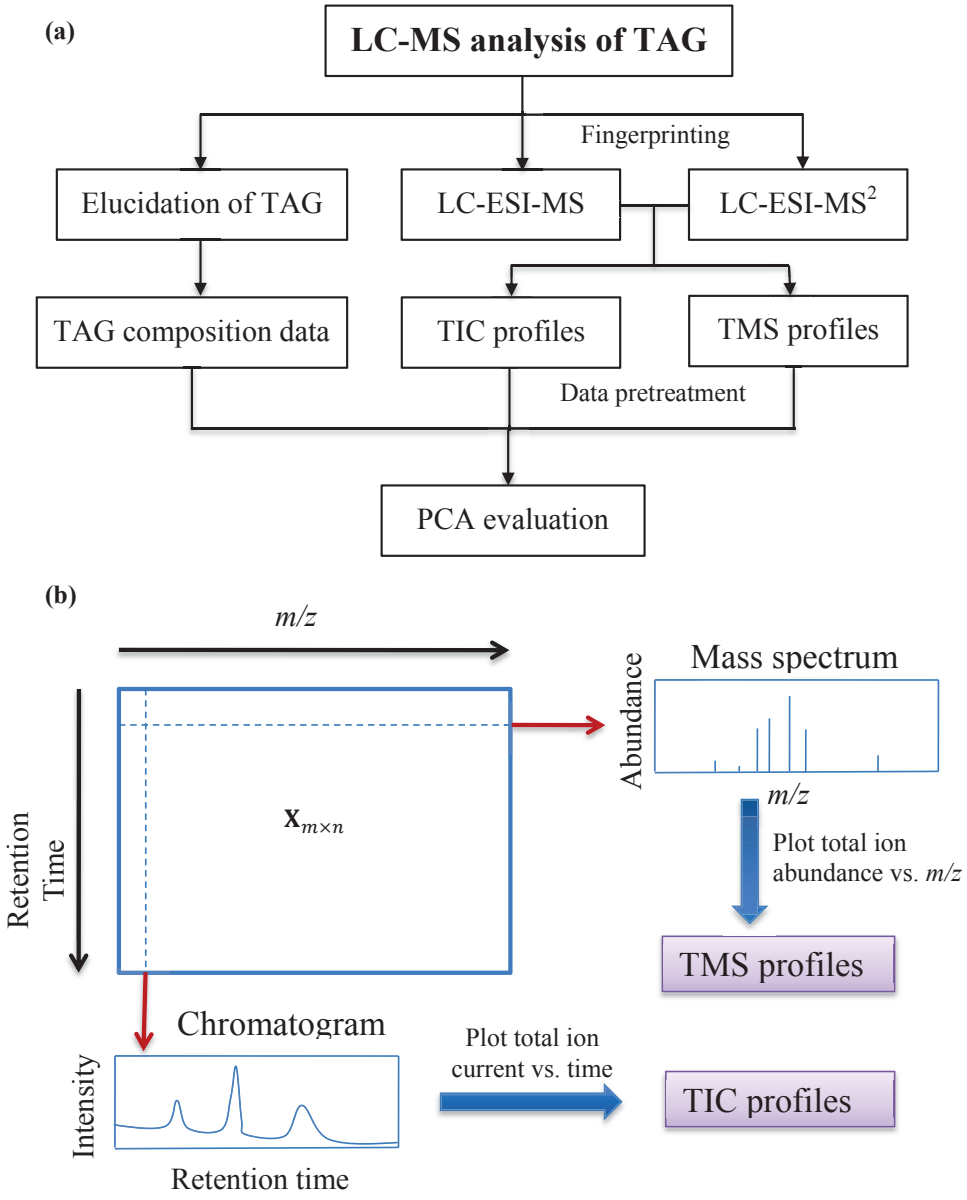


Figure 27 (a) Schematics of different strategies for obtaining data matrices for multivariate data analysis. (b) Basic features of LC-MS dataset and the generation of different TIC and TMS profiles.

The results show that the PCA study based on TAG composition enables the differentiation of marine and plant oils. However, this strategy is time-consuming since it needs elucidation of all the individual TAG species in the oil samples. PCA study of TIC profiles from LC-ESI-MS experiments does not yield any relevant pattern that allows differentiating the various oils due to severe overlapping of chromatographic peaks. Although PCA of TIC profiles from LC-ESI-MS^2 experiments enables the differentiation of marine oils from plant oils, some of the marine oils are misclassified. This demonstrates that both TIC profiles fail to represent the characteristics of TAG patterns of different oils.

The observed ions in single TMS profiles mainly corresponded to the abundant ammoniated adducts $[M+NH_4]^+$ acting as base peaks. The PCA score plot based on single TMS profiles enables basic differentiation between marine oils and plant oils. No distinction among seal oils of different qualities (*e.g.* seal refined and seal C) was observed.

The tandem TMS profiles are characterized by DAG fragments of the $[M+NH_4-R_iCOONH_4]^+$ and $[M+H]^+$ ions. Compared to single TMS profiles, the tandem TMS profiles can represent the characteristics of TAG patterns more satisfactorily. In addition, unlike the PCA study based on compositional data of TAG, tandem TMS profiles can distinguish different kinds of oils by using the various *m/z* values as the variables without the need to identify individual TAG species. The evaluation of tandem TMS profiles represents a rapid and convenient tool in differentiating and fingerprinting various marine and plant oils.

4 Phospholipids study

Phospholipids are the central components of biological membranes. Membrane phospholipids play important roles in various biological functions and processes, such as energy and signal transduction, cellular recognition, protein targeting and trafficking [120]. The analysis of phospholipids is mainly based on MS techniques such as LC-MS. However, the large amounts of data produced demand powerful computational tools capable of rapidly analyzing LC-MS data. In Paper **III**, a novel lipid analysis algorithm has been developed for the automated characterization and deconvolution of the LC-MS data of the main phospholipid species (PC and PE) in biological extracts analyzed by the class-specific PIS and NLS scanning methods, respectively. In Paper **IV**, the methodology is extended to support the interpretation of high-resolution MS and LC-MS data and to cover the analysis of major lipid classes including GPL, sphingolipids, glycerolipids *etc*.

4.1 Workflow

To analyze the lipidomic data, a lipid compound library needs to be generated first by using the built-in function in the program Chrombox D [121]. The lipid generator generates possible lipid compounds based on lists of common fatty acids, sphingoid bases and the lipid class core formula. Currently, the lipid compounds library includes fatty acids, GPL (PC, PE, PA, PS, PI), sphingolipids, glycerolipids (DAG, TAG). Next, a lipid mass spectral library with the theoretical isotope distributions can be generated based on the previously generated lipid compounds library. Parameters such as adduct types, mass resolution, mass offset and filter options should be set prior to the generation of the mass spectral library. The supported adduct types include $[M+H]^+$, $[M-H]^-$, $[M+NH_4]^+$, $[M+Na]^+$ *etc*. The NetCDF or mzXML formats of MS or LC-MS data can be directly read into Chrombox D, where the mass resolution should be the same as the generated mass spectral library. Data can be pretreated using CODA to eliminate noise and background from the data. These lipidomic data can then be resolved, identified and quantified by the algorithm. The details of this methodology can be found in Papers **III** and **IV**. If the data are LC-MS data, a chromatographic resolution can be performed to resolve the chromatograms of each identified compound. This will give additional information that can assist the identification and quantification of the compounds.

4.2 Application in the analysis of PC and PE from low-resolution LC-MS data

The performance of the algorithm has been tested by reference mixtures as well as biological samples, such as cod and mouse brain. The PC species from cod brain extract are analyzed by the algorithm. The total mass spectrum (Figure 2 in Paper **III**) is extracted from the LC-MS data acquired by PIS of *m/z* 184, where only the phosphocholine head group containing classes such as PC and sphingomyelin species are detected. The ions are mainly distributed in the mass region *m/z* 700-900, representing the masses of most common PC and sphingomyelin species. There is significant overlap of the isotopic patterns of the individual compounds in the total mass spectrum that is difficult to analyze manually, but that is easily handled by the algorithm. The algorithm can take advantage of the theoretical isotope patterns of individual compound and the resolution of the sum mass spectrum and chromatographic data, thus eliminating the need for isotope corrections. The results indicate that the main abundant PC species identified by the algorithm are PC 34:1, PC 36:1, PC 38:6 and PC 42:2, which are probably constituted by the following fatty acyl combinations, 16:0/18:1, 18:0/18:1 or 16:0/20:1, 16:0/22:6 or 18:1/20:5 and 18:1/24:1, respectively. The result from the validation shows that the original sum spectrum is accounted for with high accuracy by the spectrum calculated from the estimated amounts and theoretical spectra. The subsequent chromatographic resolution can give additional information about the identities of the compounds. The elution pattern of PC species follows the rule that the PC species having the same ECN values are basically in the same retention time region and the PC species are gradually eluted with increasing ECN values [72]. It has been shown that ether-linked PC species elute later than their diacyl counterparts on a RPLC column [74]. This can be used to distinguish the ether-linked species from the diacyl PC species.

The PE species from cod brain extract are also analyzed by algorithm. The LC-MS data is obtained by NLS of 141 Da, where only PE species that yield characteristic neutral loss of 141 Da (*i.e.*, phosphoethanolamine headgroup) are detected. The results for the PE analysis can be found in Paper **III**. The extracted total mass spectrum (Figure 4 in Paper **III**) showed that the most abundant ions are *m/z* 836, 764, 792 and 790, corresponding to the base peaks of PE 44:12, PE 38:6, PE 40:6, PE 40:7, respectively. Based on the fatty acid composition and the information from negative ionization LC-MS mode, these dominating PE have the following fatty acyl combinations: 22:6/22:6 for PE 44:12, 16:0/22:6 or 18:1/20:5 for PE 38:6,

18:0/22:6, 18:1/22:5 or 20:1/20:5 for PE 40:6 and 18:1/22:6 for PE 40:7. The chromatographic resolution is also performed afterwards. The elution pattern of diacyl PE and ether-linked PE species can be used as a diagnostic tool for the accurate identification of these species. The resolved chromatogram indicates the presence of several isobaric species, which have the same mass but different fatty acyl combinations. For example, PE 38:6 has two closely eluting chromatographic peaks, possibly consisting of 16:0/22:6 and 18:1/20:5.

4.3 Application in the analysis of phospholipids from high-resolution MS and LC-MS data

The algorithm has been extended to support the analysis of high resolution shotgun and LC-MS lipidomic data acquired by LTQ Orbitrap MS and to cover the analysis of other major lipid classes (*e.g* GPL, sphingolipids, glycerolipids *etc*). The examples for the analysis of GPL and sphingolipids from natural mixtures analyzed by high-resolution MS and LC-MS are presented and details can be found in Paper **IV**.

4.3.1 Direct infusion data

The direct infusion MS data (mzXML format) of egg PA, porcine brain PS and bovine liver PI extracts are imported into the algorithm (Figure 2 in Paper **IV**). The lipid library applied for the analyses contains all the GPL classes, which covers 2596 compounds (excluding all the isomers). The application of LSSR resolution shows that the majority of the signal is accounted for by the ions in the generated library. There are some minor ions that are not accounted for by the compounds in the library, particularly in the PI and PS samples. These ions might be interferences from other compounds in the samples or in the infusion solvent. All the validation plots show good accuracy between measured and calculated abundances, indicating that the deviation between the predicated and measured total spectrum is quite small. Some exceptions might occur for ions of very low abundance, because these ions can be easily influenced by the noise in the raw data. Cases with severe deviations and a measured value of zero usually indicate the absence of the ion in the raw data.

The analysis by the algorithm indicates that the PA sample contains almost exclusively the diacyl PA species, only one lyso-PA was observed, which might be due to the hydrolysis of the corresponding diacyl PA species. The PI sample contains one compound with an ether

bound fatty acid and all other compounds belong to the diacyl PI class. The PS sample is less pure than the other two samples. For example, a highly polar PC plasmalogen is detected in the PS sample, which can only be explained by the presence of two 22:6 fatty acids. The identity of this compound may be incorrect since no entry for the molecular formula is currently listed in the LIPID MAPS library. In addition, a few PI and PA species are observed in the PS sample, possibly due to the impurities of the sample or carryover from PI and PA samples. The analysis results obtained by the algorithm are in close accordance with other studies [122, 123] and the fatty acid composition data provided by the manufacturer [124, 125].

4.3.2 LC-MS data of brain sphingolipids

For the analysis of brain sphingolipids LC-MS data, a specific library including ceramide and sphingomyelin classes is generated. This targeted library covers 312 compounds and 238 compounds are present after filtering the isomers. Since the raw LC-MS data contain a lot of noise and interferences from other compounds, the library filter is applied, where the ions that are not present in the generated library are removed (Figure 4 in Paper **IV**). In addition, the filtering also removes several spikes or narrow peaks. The pretreatment of the LC-MS data can improve the accuracy of subsequent identification and quantification. The validation plots indicate good accuracy. Only one severe deviation is observed for 648.6 Da in the plot for ceramides, suggesting the quantification of the compound containing this ion may not be reliable.

The brain ceramide sample contains basically ceramide class with the exception of one sphingomyelin species (Figure 6 in Paper **IV**). Similarly, the brain sphingomyelin sample contains only sphingomyelin species with the exception of two ceramide species (Figure 6 in Paper **IV**). These might due to the chromatographic carryover from other samples. The brain sphingolipids contain basically saturated fatty acids (18:0, 20:0, 22:0, 24:0) and 24:1, with 18:0 as the dominating fatty acids. This is consistent with fatty acid composition reported elsewhere [126] as well as the fatty acid composition data provided by the manufacturer [127].

The LC-MS chromatograms can be resolved by using the theoretical isotope distribution of the identified compounds [75]. Although the ceramide chromatogram has more well defined peaks than the sphingomyelin chromatogram, there are some clear similarities between them (Figure 7 in Paper **IV**). For example, the retention time tends to increase with increasing chain length and decrease with increasing number of double bonds, which is expected in RPLC

system. In addition, the elution orders of the different sphingoid bases and fatty acid pairs are also similar. Some of the chromatograms have two clearly separated peaks, indicating the presence of isomers.

In summary, the described LSSR approach allows the automated analysis of direct infusion MS or LC-MS data for the lipid species. Currently, it supports the data acquired by various high or low resolution MS instrumentations such as LTQ Orbitrap MS, QqQ MS, TOF MS and ion trap MS. The integrated functions for customized generation of lipid compounds and mass spectral libraries enable the analysis of major lipid classes including GPL, sphingolipids, glycerolipids, and can be extended to support more lipid classes. The main feature of the described approach is the least squares resolution of the MS data from the calculation of theoretical mass spectra with isotope distribution, which eliminates the need for isotope correction. The described program is efficient and flexible that can be a powerful tool for lipidomics studies.

5 Application: Evaluation of effects of EPA in modulating MeHg toxicity in mice

This chapter is based on the study reported in Papers **V** and **VI**. Methylmercury (MeHg) is a ubiquitous environmental contaminant, posing a great threat to human health. One of the mechanisms of MeHg-induced toxicity is the activation of PLA_2 and the subsequent release of AA [128, 129]. Recently, the counteracting effect of *n*-3 PUFA against MeHg toxicity is receiving growing attention. However, the mechanism of *n*-3 PUFA, such as EPA, against MeHg toxicity has not yet been illustrated. Since both EPA and AA are precursors of eicosanoids, and EPA can compete with AA for the eicosanoid substrate and incorporation into cell membrane phospholipids [21, 22], we hypothesize that the supplementation of EPA may favorably modulate the MeHg toxicity through modifying the membrane phospholipid composition in mice. Papers **V** and **VI** investigate the effects of EPA in modulating the MeHg toxicity through the evaluation of phospholipid composition in mouse brain, liver and plasma.

5.1 Effects of EPA and MeHg on PC and PE composition in mice

The work of Papers **V** and **VI** is based on a two-level factorial design with four diets: high fat control diet (HF), EPA enriched diet (EPA), MeHg enriched diet (MeHg), and EPA+MeHg enriched diet (EPA+MeHg). The intact membrane PC and PE species in the four groups of mouse tissues are analyzed by LC-MS and elucidated by the LSSR algorithm. The effects of membrane PC and PE composition are subsequently evaluated by PCA and ANOVA. The results show that not only EPA, but also MeHg, are capable of modifying the PC and PE composition in the membrane phospholipid bilayer of mouse tissues, in different manners and to different extents. This is demonstrated by the LC-MS analysis of intact PC and PE species in mouse tissues followed by PCA and ANOVA.

The exploratory PCA study shows that the molecular composition of PC and PE enable the discrimination of mice fed with the different diets, indicating that these diets exhibit their particular effects on the PC and PE composition in mouse tissues. The score and loading plots of brain PC and PE are shown in Figure 28. The corresponding figures for liver and plasma can be found in Paper **VI**. To improve readability of the score and loading plots, unequal scaling of the axes has been applied. This distorts the relative placements of the individual

samples and variables. However, our interpretations deal with the signs of the scores and loadings on each component (*i.e.*, in which quadrant of the plot the samples and variable lie), and this is not altered by the scaling.

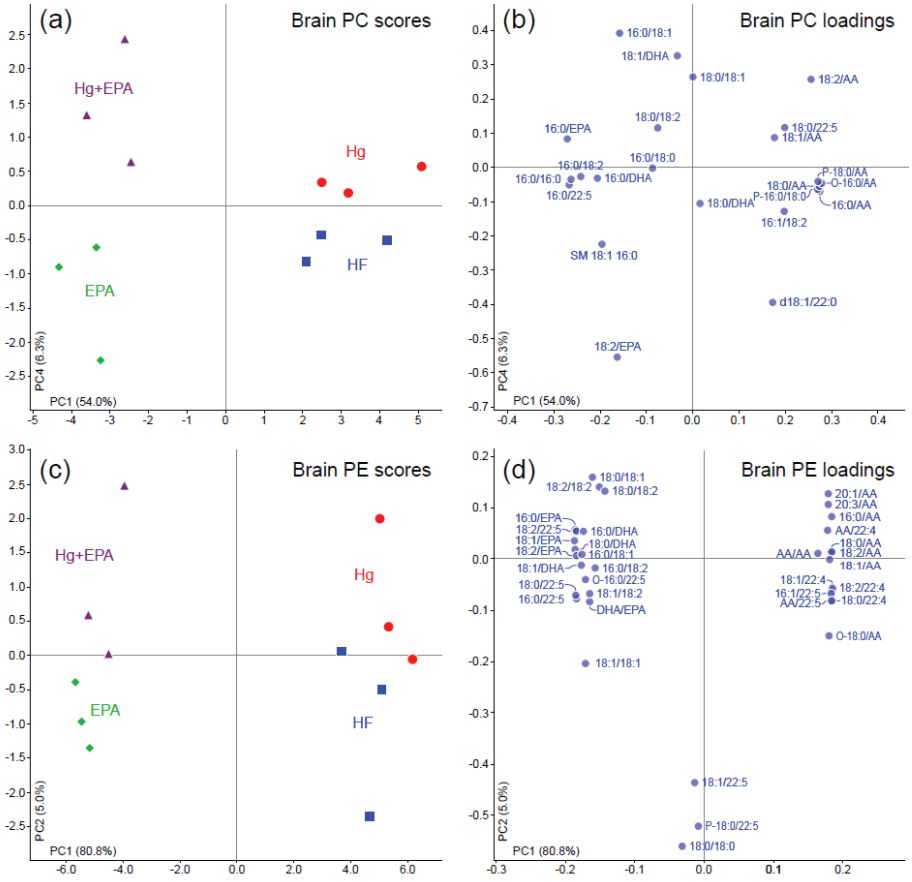


Figure 28 Scores and loadings plots of PC (a, b) and PE (c, d) profiles from mouse brain

The score plots show that EPA exerts much stronger effects on PC and PE composition than MeHg, which is also confirmed by the ANOVA. The loading plots reveal the PC and PE species that contribute to the discrimination of the different groups. The general pattern is that PC and PE species that contain *n*-3 PUFA are in the same location as the EPA-supplemented groups, while PC and PE species that contain AA are in the opposite direction. This means that EPA supplementation tends to increase in *n*-3 PUFA and decrease in AA in the measured phospholipids. For MeHg it can be seen that the majority of phospholipids containing AA are

located with the MeHg supplemented groups (above the origin) while the majority of the species that contain *n*-3 PUFA is in the opposite direction. This indicates that MeHg supplementation increases the level of AA and decreases the level of *n*-3 PUFA in the phospholipids.

The subsequent ANOVA evaluation confirms the findings from PCA and quantifies the effects of EPA on individual PC and PE species in mouse tissues. The results demonstrated that, in general, EPA supplementation results in significantly increased levels of *n*-3 PUFA containing PC and PE species, and significantly reduces levels of AA containing PC and PE species in the mouse tissues. In the case of MeHg, the results showed that MeHg supplementation exhibited significant positive effects on the AA containing PC and PE species in brain and liver, but to a much lower extent compared to EPA. In addition, MeHg has significant negative effects on *n*-3 PUFA containing PC species in liver and *n*-3 PUFA containing PE species in brain. The detailed effects of EPA and MeHg on the individual PC and PE species can be found in Papers **V** and **VI**.

5.2 Biological and toxicological analysis

Some biological and toxicological indexes are measured in order to evaluate if EPA could modify the MeHg toxicity. The analyses of PGE_2 and PGE_3 productions in plasma indicate that MeHg treated groups display higher PGE_2 and lower PGE_3, suggesting that MeHg might stimulate the production of AA and attenuate the production of *n*-3 PUFAs, while EPA treated groups show lower PGE_2 and higher PGE_3, suggesting that accumulation of *n*-3 PUFA in phospholipid species will decrease the production of AA-derived inflammatory factors. The histological analysis of cell damage and necrosis and the measurements of biochemical indexes including thiobarbituric acid reactive substances and glutathione, and activities of alanine aminotransferase and aspartate aminotransferase, demonstrate that EPA could alleviate the MeHg-induced hepatic necrosis and toxicity.

It has been established that one of the mechanisms of MeHg-induced toxicity lies is the activation of PLA_2 and the subsequent release of AA [128, 129]. Our results also show that MeHg causes specific accumulation of AA in membrane PC and PE species in mice. AA is the main substrate for synthesis of various pro-inflammatory eicosanoid mediators, thus the MeHg-triggered increases of AA-derived eicosanoids may cause excess endogenous inflammation and lead to a series of biological dysfunctions [21, 130]. EPA is an alternative

precursor of eicosanoids who can compete with AA for the eicosanoid substrate and incorporation into cell membrane phospholipids and inhibit the production of AA-derived inflammatory factors [21, 22]. Our results show that dietary EPA increases the *n*-3 PUFA levels in PC and PE species of different tissues, but decreases the AA levels in PC and PE species, suggesting EPA could decrease AA-derived pro-inflammatory factors by reducing the amount of the substrate of PLA_2. Therefore, EPA may have alleviating effects on toxicity of MeHg by inhibiting these AA-derived inflammatory factors.

6 Conclusions and future perspectives

6.1 Conclusions

The present thesis is mainly focused on the characterization of TAG in food (plant, marine oils) and phospholipids in biological systems (cod brain, mouse brain/liver/plasma) by using LC-MS and chemometrics. The main conclusions of the present study can be summarized as follows:

1) A reliable LC-ESI-MS^2 strategy for the detection and characterization of TAG species from cod liver oil was established. A computational algorithm for rapid and automatic interpretation of TAG species was developed (Paper **I**).

2) The capabilities of LC-ESI-MS and LC-ESI-MS^2 for fingerprinting and differentiating marine and plant oils based on different TAG profiles were assessed by PCA. The use of tandem mass spectral profiles was demonstrated to be the best and most feasible strategy for fingerprinting and differentiating marine and plant oils (Paper **II**).

3) A methodology has been proposed for the automated characterization, quantification and deconvolution of the main GPL species (PC and PE) in biological extracts analyzed by class-specific scanning methods in LC-MS. The described algorithm was able to resolve PC and PE species of reference mixtures, porcine brain sphingomyelin and extracts of brain lipids (Paper **III**).

4) The LSSR methodology has been further extended in two aspects: i) compatibility with high resolution MS data; ii) the analysis of other major lipid classes, such as more GPL classes, sphingolipids, and glycerolipids. The functionality of this methodology has been demonstrated by the analysis of natural lipid extracts from egg, porcine brain and bovine liver. (Papers **IV**).

5) The methodology proposed in Papers **III** and **IV** was applied to evaluate the effects of MeHg and EPA on the intact PC and PE composition in mouse brain using a two-level factorial design. The results showed that EPA led to large reductions in the levels of AA containing PC and PE species in brain, while MeHg tended to elevate the levels of AA containing PC and PE species. EPA also significantly increased the levels of *n*-3 PUFA

containing PC and PE species in brain. The results indicated that EPA may counteract the alterations of the PC and PE pattern induced by MeHg and thus alleviate the MeHg neurotoxicity in mouse brain through the inhibition of AA-derived pro-inflammatory factors (Paper **V**).

6) The work in Paper **V** was extended to mouse liver and plasma and the effects of MeHg and EPA on the PC and PE composition in liver and plasma were studied by LC-MS in conjunction with some biological and toxicological analyses. Similar to the observations in brain lipids, EPA significantly elevated the levels of PC and PE species that contain *n*-3 PUFA and reduced the levels of PC and PE species that contain AA. MeHg increased the levels of PC and PE species that contain AA, but to a lower extent. MeHg induced more PGE_2 and less PGE_3, thus increasing pro-inflammatory factors, while EPA displayed the ability to decrease the AA-derived inflammatory factors. The histological analysis of cell damage and necrosis and the measurements of biochemical indexes also indicated that MeHg induced chronic inflammatory symptoms in mice and that EPA can alleviate the MeHg-induced hepatic toxicity. In summary, EPA may have protective effects against MeHg-induced toxicity in mice due to the favorable modification of membrane phospholipid composition and the inhibition of inflammatory factors release (Paper **VI**).

6.2 Future perspectives

The work on TAG may be extended to systematic studies of regioisomeric analysis of TAG based on silver ion chromatography and the combinations of different types of mass analyzers. The use of silver ion LC can give a better resolution of the TAG compounds, where TAG can be separated according to the overall degree of unsaturation of the constituent fatty acids, and the regioisomeric TAG pairs such as AAB/ABA might be resolved. The effect of different types of mass analyzers on the mass spectra such as the relative intensities of DAG fragment ions can be evaluated and used for the regioisomeric analyses of the distribution of acyl chains on the TAG backbone. For this purpose, a broad range of standard mixtures of TAG regioisomers is needed. TAG standards that contain *n*-3 PUFAs, such as EPA and DHA, are also needed if the samples are of marine origin. In addition, for the adulterations of marine oils, a large set of samples can be included in the study using supervised pattern recognition methods, such as PLS-discriminant analysis (PLS-DA).

The LSSR methodology for the analysis of phospholipids is quite flexible and can be extended

in many aspects. At present, the methodology not only enables the analysis of the low resolution LC-MS data acquired by QqQ or ion-trap mass analyzers, but also supports the analysis of MS data acquired by high resolution Orbitrap mass spectrometry. The flexibility of the algorithm makes it possible to be compatible with the MS data acquired by different types of mass analyzers. In addition, it can be expanded to other lipid classes that are not covered by the current algorithm. If the corresponding MS/MS spectra are acquired for each precursor ion, automated structural assignment might be possible for the lipid classes with known fragmentation mechanisms, such as GPL (PC, PE, PA, PI, PS, PG), ceramides, sphingomyelins and TAG. This would give a more accurate structural elucidation of the lipids and eliminate the need of manual interpretations of MS/MS spectra. LC-MS has proved to be a powerful tool in lipidomics. On the data analytical side, there is ample room for improvement in both noise removal (*e.g.*, modifications of the CODA algorithm) and in curve resolution techniques tailor-made for LC-MS data.

References

1. Christie, W. W., and Han, X. (2012) Lipid analysis isolation, separation, identification and structural analysis of lipids. In *Lipids: their structure and occurrence.* Vol. 24 pp. 3-5, The Oily Press, Cambridge UK.
2. Ivanova, P. T., Milne, S. B., Byrne, M. O., Xiang, Y., and Brown, H. A. (2007) Glycerophospholipid identification and quantitation by electrospray ionization mass spectrometry. In *Methods in Enzymology* (Brown, H. A., ed) Vol. 432 pp. 21-57, Academic Press, London, UK.
3. Fahy, E., Cotter, D., Sud, M., and Subramaniam, S. (2011) Lipid classification, structures and tools. *BBA-Mol. Cell Biol. L.* **1811**, 637-647.
4. LIPID Metabolites and Pathways Strategy (LIPID MAPS). Available: http://www.lipidmaps.org/ [Accessed 18 May 2015].
5. Brügger, B. (2014) Lipidomics: analysis of the lipid composition of cells and subcellular organelles by electrospray ionization mass spectrometry. *Annu. Rev. Biochem.* **83**, 79-98.
6. Lipid Library website. Available: http://lipidlibrary.aocs.org [Accessed 18 May 2015].
7. Cyberlipid Center website. Available: http://www.cyberlipid.org [Accessed 18 May 2015].
8. Fahy, E., Subramaniam, S., Murphy, R. C., Nishijima, M., Raetz, C. R. H., Shimizu, T., Spener, F., van Meer, G., Wakelam, M. J. O., and Dennis, E. A. (2009) Update of the LIPID MAPS comprehensive classification system for lipids. *J. Lipid Res.* **50**, S9-S14.
9. Han, X. L., and Gross, R. W. (2005) Shotgun lipidomics: Electrospray ionization mass spectrometric analysis and quantitation of cellular lipidomes directly from crude extracts of biological samples. *Mass Spectrom Rev* **24**, 367-412.
10. Wenk, M. R. (2010) Lipidomics: New Tools and Applications. *Cell* **143**, 888-895.
11. Ekroos, K. (2012) Lipidomics perspective: from molecular lipidomics to validated clinical diagnostics. In *Lipidomics: technologies and applications* pp. 1-19, Wiley, Weinheim.
12. Quehenberger, O., Armando, A. M., and Brown, A. H. (2010) Lipidomics reveals a remarkable diversity of lipids in human plasma. *J. Lipid Res.* **51**, 3299-3305.
13. Shevchenko, A., and Simons, K. (2010) Lipidomics: coming to grips with lipid diversity. *Nat. Rev. Mol. Cell Biol.* **11**, 593-598.
14. Wymann, M. P., and Schneiter, R. (2008) Lipid signalling in disease. *Nat. Rev. Mol. Cell Bio.* **9**, 162-176.
15. Unger, R. H. (2002) Lipotoxic diseases. *Annu. Rev. Med.* **53**, 319-336.

16. Unger, R. H., and Orci, L. (2001) Diseases of liporegulation: new perspective on obesity and related disorders. *FASEB J.* **15**, 312-321.

17. O'Keefe, S. F. (2008) Food lipids: chemistry, nutrition, and biotechnology. In *Nomenclature and classification of lipids* (Akoh, C. C., and Min, D. B., eds) pp. 3-18, CRC press, Boca Raton.

18. Leonard, A. E., Pereira, S. L., Sprecher, H., and Huang, Y. S. (2004) Elongation of long-chain fatty acids. *Prog. Lipid Res.* **43**, 36-54.

19. Brenna, J. T. (2002) Efficiency of conversion of alpha-linolenic acid to long chain n-3 fatty acids in man. *Curr. Opin. Clin. Nutr. Metab. Care* **5**, 127-132.

20. Lee, K. W., and Lip, G. Y. H. (2003) The role of omega-3 fatty acids in the secondary prevention of cardiovascular disease. *QJM-Int. J. Med.* **96**, 465-480.

21. Calder, P. C. (2006) n-3 polyunsaturated fatty acids, inflammation, and inflammatory diseases. *Am J Clin Nutr* **83**, 1505s-1519s.

22. Simopoulos, A. P. (2002) Omega-3 fatty acids in inflammation and autoimmune diseases. *J. Am. Coll. Nutr.* **21**, 495-505.

23. Calder, P. C. (2001) Polyunsaturated fatty acids, inflammation, and immunity. *Lipids* **36**, 1007-1024.

24. Gil, A. (2002) Polyunsaturated fatty acids and inflammatory diseases. *Biomed. Pharmacother.* **56**, 388-396.

25. Kitajka, K., Sinclair, A. J., Weisinger, R. S., Weisinger, H. S., Mathai, M., Jayasooriya, A. P., Halver, J. E., and Puskas, L. G. (2004) Effects of dietary omega-3 polyunsaturated fatty acids on brain gene expression. *Proc. Natl. Acad. Sci. USA* **101**, 10931-10936.

26. Martinez, M., and Mougan, I. (1998) Fatty acid composition of human brain phospholipids during normal development. *J Neurochem* **71**, 2528-2533.

27. Crawford, M. A., Costeloe, K., Ghebremeskel, K., Phylactos, A., Skirvin, L., and Stacey, F. (1997) Are deficits of arachidonic and docosahexaenoic acids responsible for the neural and vascular complications of preterm babies? *Am. J. Clin. Nutr.* **66**, 1032S-1041S.

28. Lauritzen, L., Hansen, H. S., Jorgensen, M. H., and Michaelsen, K. F. (2001) The essentiality of long chain n-3 fatty acids in relation to development and function of the brain and retina. *Prog Lipid Res* **40**, 1-94.

29. Spector, A. A. (1999) Essentiality of fatty acids. *Lipids* **34**, S1-S3.

30. Jump, D. B. (2002) The biochemistry of n-3 polyunsaturated fatty acids. *J. Biol. Chem.* **277**, 8755-8758.

31. Bang, H. O., Dyerberg, J., and Sinclair, H. M. (1980) The composition of the eskimo food in northwestern greenland. *Am. J. Clin. Nutr.* **33**, 2657-2661.

32. Dyerberg, J., Bang, H. O., Stoffersen, E., Moncada, S., and Vane, J. R. (1978) Eicosapentaenoic acid and prevention of thrombosis and atherosclerosis? *Lancet* **2**, 117-119.

33. He, K., Song, Y. Q., Daviglus, M. L., Liu, K., Van Horn, L., Dyer, A. R., and Greenland, P. (2004) Accumulated evidence on fish consumption and coronary heart disease mortality - A meta-analysis of cohort studies. *Circulation* **109**, 2705-2711.

34. He, K., Song, Y. Q., Daviglus, M. L., Liu, K., Van Horn, L., Dyer, A. R., Goldbourt, U., and Greenland, P. (2004) Fish consumption and incidence of stroke - A meta-analysis of cohort studies. *Stroke* **35**, 1538-1542.

35. Wang, C., Harris, W. S., Chung, M., Lichtenstein, A. H., Balk, E. M., Kupelnick, B., Jordan, H. S., and Lau, J. (2006) n-3 Fatty acids from fish or fish-oil supplements, but not alpha-linolenic acid, benefit cardiovascular disease outcomes in primary- and secondary-prevention studies: a systematic review. *Am. J. Clin. Nutr.* **84**, 5-17.

36. Simopoulos, A. P. (1991) Omega-3 fatty acids in health and disease and in growth and development. *Am. J. Clin. Nutr.* **54**, 438-463.

37. Connor, W. E. (2000) Importance of *n*-3 fatty acids in health and disease. *Am. J. Clin. Nutr.* **71**, S171-S175.

38. Venugopal, V. (2009) Marine products for healthcare: Functional and bioactive nutraceutical compounds from the ocean. In *Polyunsaturated fatty acids and their therapeutic functions* pp. 143-178, CRC Press, Boca Raton.

39. Ruxton, C. H. S., Reed, S. C., Simpson, M. J. A. and Millington, K. J. (2004) The health benefits of omega-3 polyunsaturated fatty acids: a review of the evidence. *J. Hum. Nutr. Dietet.* **17**, 449-459.

40. Simopoulos, A. P. (1999) Essential fatty acids in health and chronic disease. *Am. J. Clin. Nutr.* **70**, 560s-569s.

41. Simopoulos, A. P. (2008) The importance of the omega-6/omega-3 fatty acid ratio in cardiovascular disease and other chronic diseases. *Exp. Biol. Med.* **233**, 674-688.

42. Buchgraber, M., Ullberth, F., Emons, H., and Anklam, E. (2004) Triacylglycerol profiling by using chromatographic techniques. *Eur. J. Lipid Sci. Technol.* **106**, 621-648.

43. IUPAC-IUB Commission on biochemical nomenclature (1967) The nomenclature of lipids. *J. Lipid Res.* **8**, 523-528.

44. Wasta, Z., and Mjøs, S. A. (2013) A database of chromatographic properties and mass spectra of fatty acid methyl esters from omega-3 products. *J Chromatogr A* **1299**, 94-102.

45. Hamosh, M., and Scow, R. O. (1973) Lingual lipase and its role in digestion of dietary lipid. *J. Clin. Invest.* **52**, 88-95.

46. Hamosh, M. (1979) Review: Fat digestion in the newborn - Role of lingual lipase and pre-duodenal digestion. *Pediatr. Res.* **13**, 615-622.

47. Mu, H. L., and Hoy, C. E. (2004) The digestion of dietary triacylglycerols. *Prog. Lipid Res.* **43**, 105-133.

48. Carey, M. C., Small, D. M., and Bliss, C. M. (1983) Lipid digestion and absorption. *Annu. Rev. Physiol.* **45**, 651-677.

49. Kubow, S. (1996) The influence of positional distribution of fatty acids in native, interesterified and structure-specific lipids on lipoprotein metabolism and atherogenesis. *J. Nutr. Biochem.* **7**, 530-541.

50. McNamara, D. J. (1992) Dietary fatty acids, lipoproteins and cardiovascular disease. *Adv. Food Nutr. Res.* **6**, 254-353.

51. Mattson, F. H., and Volpenhein, R. A. (1964) The digestion and absorption of triglycerides. *J. Biol. Chem.* **239**, 2772-2777.

52. Hashim, S. A., and Babayan, V. K. (1978) Studies in man of partially absorbed dietary fats. *Am. J. Clin. Nutr.* **31**, S273-S276.

53. Ramirez, M., Amate, L., and Gil, A. (2001) Absorption and distribution of dietary fatty acids from different sources. *Early Hum. Dev.* **65**, S95-S101.

54. Christensen, M. S., Hoy, C. E., Becker, C. C., and Redgrave, T. G. (1995) Intestinal-absorption and lymphatic transport of eicosapentaenoic (EPA), docosahexaenoic (DHA), and decanoic acids - dependence on intramolecular triacylglycerol structure. *Am. J. Clin. Nutr.* **61**, 56-61.

55. Cullis, P. R., Fenske, D. B., and Hope, M. J. (1996) Physical properties and functional roles of lipids in membranes. In *New comprehensive biochemistry* (Dennis, E. V., and Jean, E. V., eds) Vol. 31 pp. 1-33, Elsevier, Amsterdam.

56. Weihrauch, J. L., and Son, Y. S. (1983) The phospholipid content of foods. *J Am Oil Chem Soc* **60**, 1971-1978.

57. Cohn, J. S., Kamili, A., Wat, E., Chung, R. W. S., and Tandy, S. (2010) Dietary phospholipids and intestinal cholesterol absorption. *Nutrients* **2**, 116-127.

58. van den Bosch, H., Postema, N. M., de Haas, G. H., and van Deenen, L. L. (1965) On the positional specificity of phospholipase A from pancreas. *Biochim. Biophys. Acta.* **98**, 657-659.

59. Borgström, B. (1976) Phospholipid absorption. In *Lipid absorption: biochemical and clinical aspects* (Rommel, K., Goebell, H., and Böhmer, R., eds) pp. 65-72, Springer, Lancaster, UK.

60. Spener, F., Lagarde, M., Géloên, A., and Record, M. (2003) Editorial: What is lipidomics? *Eur. J. Lipid Sci. Tech.* **105**, 481-482.

61. Han, X., and Gross, R. W. (2003) Global analyses of cellular lipidomes directly from crude extracts of biological samples by ESI mass spectrometry: a bridge to lipidomics. *J. Lipid Res.* **44**, 1071-1079.

62. Li, M., Yang, L., Bai, Y., and Liu, H. (2014) Analytical methods in lipidomics and their applications. *Anal. Chem.* **86**, 161-175.

63. Hyötyläinen, T., Bondia-Pons, I., and Orešič, M. (2013) Lipidomics in nutrition and food research. *Mol. Nutr. Food Res.* **57**, 1306-1318.

64. Astarita, G., Ahmed, F., and Piomelli, D. (2009) Lipidomic analysis of biological samples by liquid chromatography coupled to mass spectrometry. In *Lipidomics* (Armstrong, D., ed) Vol. 579 pp. 201-219, Humana Press, New York.

65. Murphy, S. A., and Nicolaou, A. (2013) Lipidomics applications in health, disease and nutrition research. *Mol. Nutr. Food Res.* **57**, 1336-1346.

66. Ackermann, B. L., Hale, J. E., and Duffin, K. L. (2006) The role of mass spectrometry in biomarker discovery and measurement. *Curr. Drug Metab.* **7**, 525-539.

67. Milne, S., Ivanova, P., Forrester, J., and Alex Brown, H. (2006) Lipidomics: An analysis of cellular lipids by ESI-MS. *Methods* **39**, 92-103.

68. Moreau, R. A. (2006) An overview of modern mass spectrometry methods in the toolbox of lipid chemists and biochemists. In *Lipid analysis and lipidomics: new techniques and applications* (Mossoba;, M. M., Kramer;, J. K. G., Brenna;, J. T., and McDonald, R. E., eds) pp. 29-46, AOCS Press, Champaign, Illinois.

69. Jung, H. R., Sylvanne, T., Koistinen, K. M., Tarasov, K., Kauhanen, D., and Ekroos, K. (2011) High throughput quantitative molecular lipidomics. *Biochim. Biophys. Acta.* **1811**, 925-934.

70. Brouwers, J. F. (2011) Liquid chromatographic-mass spectrometric analysis of phospholipids. Chromatography, ionization and quantification. *BBA-Mol. Cell Biol. L.* **1811**, 763-775.

71. Shukla, V. K. (1988) Recent advances in the high performance liquid chromatography of lipids. *Prog Lipid Res* **27**, 5-38.

72. Smith, M., and Jungalwala, F. B. (1981) Reversed-phase high-performance liquid chromatography of phosphatidylcholine - a simple method for determining relative hydrophobic interaction of various molecular-species. *J. Lipid Res.* **22**, 697-704.

73. Plattner, R. D., Spencer, G. F., and Kleiman, R. (1977) Triglyceride separation by reverse phase high-performance liquid-chromatography. *J Am Oil Chem Soc* **54**, 511-515.

74. Brouwers, J. F., Vernooij, E. A., Tielens, A. G., and van Golde, L. M. (1999) Rapid separation and identification of phosphatidylethanolamine molecular species. *J. Lipid Res.* **40**, 164-169.

75. Zeng, Y. X., Mjøs, S. A., Meier, S., Lin, C. C., and Vadla, R. (2013) Least squares spectral resolution of liquid chromatography-mass spectrometry data of glycerophospholipids. *J Chromatogr A* **1280**, 23-34.

76. Kim, H. Y., Wang, T. C. L., and Ma, Y. C. (1994) Liquid-chromatography mass-spectrometry of phospholipids using electrospray-ionization. *Anal. Chem.* **66**, 3977-3982.

77. Cole, R. B. (2000) Some tenets pertaining to electrospray ionization mass spectrometry. *J Mass Spectrom.* **35**, 763-772.

78. Han, X., and Gross, R. W. (2007) Global cellular lipidome analyses by shotgun lipidomics using multidimensional mass spectrometry. In *Lipid analysis and lipidomics: New techniques and applications* (Mossoba, M. M., ed) pp. 51-58, AOCS Press, Champaign, Illinois.

79. Han, X. (2010) Applications of ESI and MALDI to lipid analysis. In *Electrospray and MALDI mass spectrometry: fundamentals, instrumentation, practicalities, and biological applications* (Cole, R. B., ed) pp. 779-781, Wiley, New Jersey.

80. Han, X. L., and Gross, R. W. (1995) Structural determination of picomole amounts of phospholipids via electrospray ionization tandem mass spectrometry. *J. Am. Soc. Mass Spectrom.* **6**, 1202-1210.

81. Zeng, Y. X., Araujo, P., Du, Z. Y., Nguyen, T. T., Froyland, L., and Grung, B. (2010) Elucidation of triacylglycerols in cod liver oil by liquid chromatography electrospray tandem ion-trap mass spectrometry. *Talanta* **82**, 1261-1270.

82. Han, X. L., and Gross, R. W. (1994) Electrospray-ionization mass spectroscopic analysis of human erythrocyte plasma-membrane phospholipids. *P Natl Acad Sci USA* **91**, 10635-10639.

83. Kerwin, J. L., Tuininga, A. R., and Ericsson, L. H. (1994) Identification of molecular-species of glycerophospholipids and sphingomyelin using electrospray mass-spectrometry. *J Lipid Res* **35**, 1102-1114.

84. Quattro LC user's guide. In *MassLynx NT User's Guide* pp. 18-22, Micromass UK Limited, Wythenshawe.

85. Ivanova, P. T., Milne, S. B., Myers, D. S., and Brown, H. A. (2009) Lipidomics: a mass spectrometry based systems level analysis of cellular lipids. *Curr. Opin. Chem. Biol.* **13**, 526-531.

86. Pulfer, M., and Murphy, R. C. (2003) Electrospray mass spectrometry of phospholipids. *Mass Spectrom. Rev.* **22**, 332-364.

87. Retra, K., Bleijerveld, O. B., van Gestel, R. A., Tielens, A. G., van Hellemond, J. J., and Brouwers, J. F. (2008) A simple and universal method for the separation and identification of phospholipid molecular species. *Rapid Commun. Mass Spectrom.* **22**, 1853-1862.

88. Demartini, D. R. (2013) A short overview of the components in mass spectrometry instrumentation for proteomics analyses. In *Tandem mass spectrometry - molecular characterization* (Coelho, A. V., ed), InTech. DOI: 10.5772/54484. ISBN: 978-953-51-1136-8.

89. Cheng, C., Gross, M. L., and Pittenauer, E. (1998) Complete structural elucidation of triacylglycerols by tandem sector mass spectrometry. *Anal. Chem.* **70**, 4417-4426.

90. Hvattum, E. (2001) Analysis of triacylglycerols with non-aqueous reversed phase liquid chromatography and positive ion electrospray tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* **15**, 187-190.

91. Marzilli, L., Fay, L., Dionisi, F., and Vouros, P. (2003) Structural characterization of triacylglycerols using electrospray ionization-MS^n ion-trap MS. *J. Am. Oil Chem. Soc.* **80**, 195-202.

92. Smith, R. M. (2004) *Understanding mass spectra: A basic approach*, Wiley, Hoboken, New Jersey.

93. Haimi, P., Uphoff, A., Hermansson, M., and Somerharju, P. (2006) Software tools for analysis of mass spectrometric lipidome data. *Anal. Chem.* **78**, 8324-8331.

94. Hartler, J., Trotzmuller, M., Chitraju, C., Spener, F., Kofeler, H. C., and Thallinger, G. G. (2011) Lipid Data Analyzer: unattended identification and quantitation of lipids in LC-MS data. *Bioinformatics* **27**, 572-577.

95. Song, H., Hsu, F. F., Ladenson, J., and Turk, J. (2007) Algorithm for processing raw mass spectrometric data to identify and quantitate complex lipid molecular species in mixtures by data-dependent scanning and fragment ion database searching. *J. Am. Soc. Mass Spectrom.* **18**, 1848-1858.

96. Murray, K. K., Boyd, R. K., Eberlin, M. N., Langley, G. J., Li, L., and Naito, Y. (2013) Definitions of terms relating to mass spectrometry (IUPAC Recommendations 2013). *Pure Appl Chem* **85**, 1515-1609.

97. Marshall, A. G., and Hendrickson, C. L. (2008) High resolution mass spectrometers. *Annu Rev Anal Chem* **1**, 579-599.

98. Kind, T., and Fiehn, O. (2006) Metabolomic database annotations via query of elemental compositions: Mass accuracy is insufficient even at less than 1 ppm. *BMC Bioinformatics* **7**, 234.

99. Perry, R. H., Cooks, R. G., and Noll, R. J. (2008) Orbitrap mass spectrometry: instrumentation, ion motion and applications. *Mass Spectrom Rev.* **27**, 661-699.

100. Marshall, A. G. (2000) Milestones in Fourier transform ion cyclotron resonance mass spectrometry technique development. *Int J Mass Spectrom* **200**, 331-356.

101. Windig, W., Phalp, J. M., and Payne, A. W. (1996) A noise and background reduction method for component detection in liquid chromatography/mass spectrometry. *Anal. Chem.* **68**, 3602-3606.

102. Fredriksson, M., Petersson, P, Magnus, J.K., Bengt-Olof, A., Bylund D. (2007) An objective comparison of pre-processing methods for enhancement of liquid chromatography–mass spectrometry data. *J Chromatogr A* **1172**, 135–150.

103. Karpievitch, Y. V., Taverner, T., Adkins, J. N., Callister, S. J., Anderson, G. A., Smith, R. D., and Dabney, A. R. (2009) Normalization of peak intensities in bottom-up MS-based proteomics using singular value decomposition. *Bioinformatics* **25**, 2573-2580.

104. Muddiman, D. C., Rockwood, A.L., Gao, Q.Y., Severs J.C., Udseth H.R., Smith R.D., Proctor A. (1995) Application of sequential paired covariance to capillary electrophoresis electrospray ionization time-of-flight mass spectrometry: unraveling the signal from the noise in the electropherogram. *Anal. Chem.* **67**, 4371-4375.

105. Muddiman, D. C., Huang, B.M., Anderson G.A., Rockwood A., Hofstadler S.A. (1997) Application of sequential paired covariance to liquid chromatography-mass spectrometry data: Enhancements in both the signal-to-noise ratio and the resolution of analyte peaks in the chromatogram. *J Chromatogr A* **771**, 1-7.

106. Fleming, C. M., Kowalskia, B.R., Apffelb, A., Hancockb W.S. (1999) Windowed mass selection method: a new data processing algorithm for liquid chromatography–mass spectrometry data. *J Chromatogr A* **849**, 71-85.

107. Windig, W., Smith, W.F., Nichols W.F. (2001) Fast interpretation of complex LC/MS data using chemometrics. *Anal. Chim. Acta* **446**, 467-476.

108. Williams, A., Lee, M. S., and Lashin, V. (2001) An integrated desktop mass spectrometry processing and molecular structure management system. *Spectroscopy* **16**, 38-49.

109. Govorukhina, N. I., Reijmers, T. H., Nyangoma, S. O., van der Zee, A. G. J., Jansen, R. C., and Bischoff, R. (2006) Analysis of human serum by liquid chromatography-mass spectrometry: Improved sample preparation and data analysis. *J Chromatogr A* **1120**, 142-150.

110. Williams, J. P., Khan, M. U., and Wong, D. (1995) A simple technique for the analysis of positional distribution of fatty acids on diacylglycerols and triacylglycerols using lipase and phospholipase A_2. *J. Lipid Res.* **36**, 1407-1412.

111. Wold, S., Esbensen, K., and Geladi, P. (1987) Principal component analysis. *Chemometr. Intell. Lab. Syst.* **2**, 37-52.

112. Wenk, M. R. (2005) The emerging field of lipidomics. *Nat. Rev. Drug Discov.* **4**, 594-610.

113. Schwudke, D., Oegema, J., Burton, L., Entchev, E., Hannich, J. T., Ejsing, C. S., Kurzchalia, T., and Shevchenko, A. (2006) Lipid profiling by multiple precursor and neutral loss scanning driven by the data-dependent acquisition. *Anal. Chem.* **78**, 585-595.

114. Shimokawa, H., and Vanhoutte, P. M. (1988) Dietary cod-Liver oil improves endothelium-dependent responses in hypercholesterolemic and atherosclerotic porcine coronary-arteries. *Circulation* **78**, 1421-1430.

115. Osterud, B., Elvevoll, E., Barstad, H., Brox, J., Halvorsen, H., Lia, K., Olsen, J. O., Olsen, R. L., Sissener, C., Rekdal, O., and Vognild, E. (1995) Effect of marine oils supplementation on coagulation and cellular activation in whole-blood. *Lipids* **30**, 1111-1118.

116. Seierstad, S. L., Seljeflot, I., Johansen, O., Hansen, R., Haugen, M., Rosenlund, G., Froyland, L., and Arnesen, H. (2005) Dietary intake of differently fed salmon; the influence on markers of human atherosclerosis. *Eur. J. Clin. Invest.* **35**, 52-59.

117. Brunborg, L. A., Madland, T. M., Lind, R. A., Arslan, G., Berstad, A., and Froyland, L. (2008) Effects of short-term oral administration of dietary marine oils in patients with inflammatory bowel disease and joint pain: A pilot study comparing seal oil and cod liver oil. *Clin. Nutr.* **27**, 614-622.

118. Galarraga, B., Ho, M., Youssef, H. M., Hill, A., McMahon, H., Hall, C., Ogston, S., Nuki, G., and Belch, J. J. F. (2008) Cod liver oil (*n*-3 fatty acids) as an non-steroidal anti-inflammatory drug sparing agent in rheumatoid arthritis. *Rheumatology* **47**, 665-669.

119. Zeng, Y., Araujo, P., Grung, B., and Zhang, L. (2011) Evaluation of different fingerprinting strategies for differentiating marine oils by liquid chromatography ion-trap mass spectrometry and chemometrics. *Analyst* **136**, 1507-1514.

120. Dowhan, W. (1997) Molecular basis for membrane phospholipid diversity: Why are there so many lipids? *Annu. Rev. Biochem.* **66**, 199-232.

121. Chrombox D. Available: http://www.chrombox.org/ [Accessed 18 May 2015].

122. Shadyro, O., Yurkova, I., Kisel, M., Brede, O., and Arnhold, J. (2004) Formation of phosphatidic acid, ceramide, and diglyceride on radiolysis of lipids: identification by MALDI-TOF mass spectrometry. *Free Radic Biol Med* **36**, 1612-1624.

123. Kurvinen, J. P., Kuksis, A., Sinclair, A. J., Abedin, L., and Kallio, H. (2000) The effect of low alpha-linolenic acid diet on glycerophospholipid molecular species in guinea pig brain. *Lipids* **35**, 1001-1009.

124. Avanti Polar Lipids. Fatty acid distribution of egg phosphatidic acid. Available: http://www.avantilipids.com/index.php?option=com_content&view=article&id=306&Itemid=221&catnumber=840101 [Accessed 18 May 2015].

125. Avanti Polar Lipids. Fatty acid distribution of bovine liver phosphatidylinositol. Available: http://www.avantilipids.com/index.php?option=com_content&view=article&id=394&Itemid=242&catnumber=840042 [Accessed 18 May 2015].

126. O'Brien, J. S., and Rouser, G. (1964) The fatty acid composition of brain sphingolipids: sphingomyelin, ceramide, cerebroside, and cerebroside sulfate. *J Lipid Res* **5**, 339-342.

127. Avanti Polar Lipids. Fatty acid distribution of porcine brain sphingomyelin. Available: http://www.avantilipids.com/index.php?option=com_content&view=article&id=436&Itemid=277&catnumber=860062 [Accessed 18 May 2015].

128. Verity, M. A., Sarafian, T., Pacifici, E. H. K., and Sevanian, A. (1994) Phospholipase A_2 stimulation by methyl mercury in neuron culture. *J Neurochem* **62**, 705-714.

129. Shanker, G., Mutkus, L. A., Walker, S. J., and Aschner, M. (2002) Methylmercury enhances arachidonic acid release and cytosolic phospholipase A_2 expression in primary cultures of neonatal astrocytes. *Mol Brain Res* **106**, 1-11.

130. Ricciotti, E., and FitzGerald, G. A. (2011) Prostaglandins and inflammation. *Arterioscl. Throm. Vas.* **31**, 986-1000.

I

Paper I

Elucidation of triacylglycerols in cod liver oil by liquid chromatography electrospray tandem ion trap mass spectrometry

Y.X. Zeng, P. Araujo, Z.Y. Du, T.T. Nguyen, L. Frøyland, B. Grung, *Talanta* 82: 1261–1270 (2010).

Talanta 82 (2010) 1261–1270

Contents lists available at ScienceDirect

Talanta

journal homepage: www.elsevier.com/locate/talanta

Elucidation of triacylglycerols in cod liver oil by liquid chromatography electrospray tandem ion-trap mass spectrometry

Ying-Xu Zeng[a,b,c], Pedro Araujo[a,*], Zhen-Yu Du[a], Thu-Thao Nguyen[a], Livar Frøyland[a], Bjørn Grung[b]

[a] National Institute of Nutrition and Seafood Research (NIFES), PO Box 2029, Nordnes, N-5817 Bergen, Norway
[b] Department of Chemistry, University of Bergen, N-5009 Bergen, Norway
[c] Faculty of Sciences and Technology, University of Algarve, Campus de Gambelas, 8005-139 Faro, Portugal

ARTICLE INFO

Article history:
Received 22 March 2010
Received in revised form 4 June 2010
Accepted 28 June 2010
Available online 24 July 2010

Keywords:
Cod liver oil
Triacylglycerols
Fatty acids
Liquid chromatography electrospray
tandem mass spectrometry
Algorithm

ABSTRACT

Though liquid chromatography electrospray tandem mass spectrometry (LC–ESI-MS^2) has been widely used in the structural elucidation of triacylglycerols (TAG) in vegetable oils, its potentiality for the identification of TAG molecules in omega-3 rich oils remains unexplored till date. Hence, this article investigates the applicability of LC–ESI-MS^2 for the structural characterization of naturally occurring TAG in cod liver oil without the TAG fractionation during the sample preparation. A computational algorithm was developed to automatically interpret the mass spectra and elucidate the TAG structures respectively. The results were compared against the lipase benchmark method. A principal component analysis study revealed that it is possible to discriminate genuine from adulterated cod liver oil.

1. Introduction

Cod liver oil has attracted extensive interests due to the scientific evidence and consumer awareness of its nutritional advantages attributed to the abundant content of omega-3 (ω-3) fatty acids (FAs) such as eicosapentaenoic acid (20:5*n*-3; EPA) and docosahexaenoic acid (22:6*n*-3; DHA) present in the form of triacylglycerols (TAG) [1–5].

Cod liver oil mainly contains TAG consisting of various esterified FAs at the three available stereospecific positions (*sn*-1, *sn*-2 and *sn*-3) of a glycerol molecule. Analysis of TAG in ω-3 rich oils is quite challenging due to the presence of a large number of positional and structural TAG isomers with very similar chemical and physical properties. Traditional chemical/enzymatic hydrolysis methods (Grignard reagent or lipases) [6–11] and sophisticated high resolution nuclear magnetic resonance spectrometry methods (^{13}C NMR or ^{1}H NMR) [12–14] have been used for the stereospecific analysis of TAG in ω-3 rich oils. In general, the titles of published articles on the analysis of TAG in ω-3 rich oils by these approaches seem to imply the elucidation of TAG structures. However, a close inspection of these articles demonstrated that they cannot provide any information regarding the structural elucidation of intact TAG not to mention positional isomers. Instead, they are mainly concerned with the quantification of the "total amount" of individual FAs at *sn*-1, *sn*-2 and *sn*-3 spatial positions. For instance, chemical hydrolysis [11], ^{13}C NMR [13] and ^{1}H NMR [14] have been implemented in the analysis of different fish oils (e.g. cod liver oil) for determining the amounts of esterified FAs at *sn*-1, *sn*-2 and *sn*-3, however the exact position of the various FAs on the backbone of the glycerol molecules was not determined. Traditional hydrolysis methods are characterized by laborious and time-consuming sample preparation protocols such as the cleavage of one or two FAs from intact TAG in order to produce the monoacylglycerols (MAG) or diacylglycerols (DAG); multiple extractions of the various free FAs, MAG or DAG; methylation of the various fractions prior to gas chromatography (GC); derivatization of the MAG and DAG fractions prior to high-performance liquid chromatography (HPLC) [6–11]. In addition, these steps are not always applicable since they are often accompanied by problems such as restrictions due to the intrinsic characteristics of the lipase, inaccuracies due to the incidence of acyl migration and hydrolysis selectivity [15–18]. Sophisticated NMR methods are affected by the presence of strongly overlapping signals, and the effect on chemical shift of the neighboring chains which in turn affect the carbonyl region by preventing the extraction of any qualitative or quantitative information in this region and rendering the C2 region (signal relative

* Corresponding author. Tel.: +47 95285039; fax: +47 55905299.
E-mail address: Pedro.Araujo@nifes.no (P. Araujo).

0039-9140/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.talanta.2010.06.055

to *sn*-2 position) unsuitable for the analysis of FAs composition [19,20].

The structural elucidation of the exact positioning of the various FAs on the glycerol molecules is essential for understanding the physiology of food processing. It has been demonstrated that FAs at *sn*-1 and *sn*-3 of the TAG are hydrolyzed during digestion and absorption of dietary oils while FAs at the *sn*-2 position remain intact [21]. Numerous studies have also shown that the positioning of FAs on the backbone of TAG molecules could affect many lipid properties such as physical and nutritional properties, oxidative stability, lipid absorption, metabolism and atherogenesis [21–24]. In addition, the determination of the stereospecific positioning of FAs on TAG (especially those at *sn*-2) could help to evaluate the quality and authenticity of nutritional ω-3 rich oils such as cod liver oil. Nowadays, the worldwide growing popularity of edible fish and ω-3 rich oils is acknowledged in rich and poor nations where they are making newspaper headlines due to their associated health benefits and also their adulteration [25,26]. For instance, the newspaper with the widest circulation in United States has recently regarded fish as the most frequently adulterated food in America [25]. In addition, it should be mentioned that the importance of developing techniques aiming at detecting adulteration of fish oils has been emphasized since the late 19th early 20th century when a great scarcity of cod liver oil accompanied by famine prices of the market brought about adulteration of genuine cod liver oil with low-grade shark oil [27,28].

For these reasons, national and international organisations have encouraged and supported the development of reliable methods for the analysis of ω-3 rich oils, such as cod liver oil, not only with the capacity to characterize quantitatively the FAs on the glycerol backbone but also to elucidate qualitatively the structures of intact TAG. The combination of these quantitative and qualitative results will assist in gaining a better knowledge of their various properties, nutritional values, commercial quality and the involvement of specific chemical structures in different human and animal physiological processes [29,30].

Several instrumental techniques such as GC, HPLC, silver-ion HPLC with mass spectrometry (MS), HPLC with fast atom bombardment-MS (FAB-MS), have been used for elucidating the structures of intact TAG in dietary ω-3 rich oils [31–33]. However, the commonly persistent limitation is the exclusive elucidation of TAG structures that can be resolved by chromatographic means and matched to commercially available TAG reference standards [33]. Such a limitation becomes a serious problem for the elucidation of TAG structures in ω-3 rich oils due the complexity of their naturally occurring TAG species. Other problems associated with these instrumental techniques are the tedious sample preparation protocols and the application of complex mathematical equations and models based on the specialized theories for identification purpose [31,32,34].

Liquid chromatography electrospray tandem MS (LC–ESI-MS^2) has been effectively used in the elucidation of TAG structures in a range of simple plant oils [35–39]. However, it is surprising the current literature on the elucidation of TAG structures in ω-3 rich oils has ignored its potentiality. The reason behind this lack of interest could be the enormous amount of time required by manual data analysis of the very complex chromatograms characteristic of ω-3 rich oils. It can be foreseen that the application of LC–ESI-MS^2 in conjunction with the automation of the interpretation process might offer a powerful means for elucidating TAG structures in cod liver oil.

The objective of the present study is to explore the capability of LC–ESI-MS^2 to identify the relative arrangement of the acyl groups on intact TAG molecules in cod liver oil. By using the basic structural features of a TAG molecule and its fragmentation mechanism, a computational algorithm is developed to assist the interpretation and prediction processes. The elucidated spatial positioning of the various acyl groups by LC–ESI-MS^2 was compared against the well-established lipase method. To our knowledge, this is the first study on structural elucidation of TAG molecules present in cod liver oil by LC–ESI-MS^2.

2. Experimental

2.1. Materials and reagents

1-Arachidin-2-Olein-3-Palmitin-glycerol (AOP), 1-Arachidin-2-Palmitin-3-Olein-glycerol (APO), 1-Palmitin-2-Arachidin-3-Olein-glycerol (PAO), 1-Arachidin-2-Linolein-3-Olein-glycerol (ALO), and 1-Palmitin-2-Olein-3-Linolein-glycerol (POL) were from Larodan Fine Chemicals (Malmö, Sweden). 1,2,3-α-Linolenoyl-glycerol (LnLnLn) and butylated hydroxytoluene (BHT) were from Sigma–Aldrich Corporation (St. Louis, MO, USA). Mixtures of the TAG standards were prepared in a chloroform:methanol (2:1, v/v) solution. Cod liver oil was from Peter Möller (Lysaker, Norway). Linseed and rapeseed oils were from Kinsarvik Naturkost (Bergen, Norway), soy oil was from Mills DA (Sofienberg, Norway) and seal oil was from Rieber Skinn A/S (Bergen, Norway). All solvents were HPLC grade. Lipase from *Rhizopus arrhizus* was obtained from Sigma–Aldrich (Schnelldorf, Germany). Fatty acid methyl ester (FAME) pure standards and also model mixture standards 2A and 2B ($C_{18:0}$, $C_{18:1n\text{-}9}$, $C_{18:2n\text{-}6}$, $C_{18:3n\text{-}3}$, $C_{20:4n\text{-}6}$), 3A ($C_{18:2n\text{-}6}$, $C_{18:3n\text{-}3}$, $C_{20:4n\text{-}6}$, $C_{22:6n\text{-}3}$), 4A ($C_{6:0}$, $C_{8:0}$, $C_{10:0}$, $C_{12:0}$, $C_{14:0}$), 6A ($C_{16:0}$, $C_{18:0}$, $C_{20:0}$, $C_{22:0}$, $C_{24:0}$), 7A ($C_{16:1n\text{-}7}$, $C_{16:1n\text{-}9}$, $C_{20:1n\text{-}9}$, $C_{22:1n\text{-}11}$, $C_{24:1n\text{-}9}$) and 14A ($C_{13:0}$, $C_{15:0}$, $C_{17:0}$, $C_{19:0}$, $C_{21:0}$) were purchased from Nu-Chek Prep (Elysian, MN). Nonadecanoic acid methyl ester ($C_{19:0}$) internal standard and formic acid were from Fluka (Buchs, Switzerland).

2.2. Sample protocols

2.2.1. Lipase method

The protocol was slightly modified from the procedure described elsewhere [40]. Briefly, 1 ml of Tris–HCl buffer (40 mM, pH 7.2) containing 50 mM of sodium borate was added to a nitrogen-dried oil sample (1 ml) and the mixture sonicated for 10 min. 60 μl of lipase (150 units) were added to the sonicated mixture and incubated at 22 °C for up to 60 min with continuous shaking. The reaction was stopped by adding 0.8 ml of acetic acid (0.1 M) and the total lipids exacted by adding 3 ml of chloroform/methanol (2:1, v/v). The lipid solution was divided into two equal portions (I and II), dried under nitrogen and methylated for 30 and 2 min at room temperature and in a microwave oven by using 1 ml methanolic solutions of NaOH (0.1 N) and HCl (0.2 N) for portion I and II respectively. The FAME in each methylation reactor were extracted into hexane after the addition of 0.2 ml of water to the reaction mixture. The hexane extracts of the NaOH reaction were washed once with water to remove any trace of NaOH before drying under nitrogen. The dried FAME extracts were redissolved in hexane and analyzed by GC. The FAME were estimated quantitatively by using $C_{19:0}$ internal standard. The lipase method was also applied to the TAG standards dissolved in chloroform:methanol (2:1, v/v). It must be mentioned that the acidic reaction allows the methylation of both DAG and FAs generated by the lipase procedure, while the basic reaction allows exclusively the methylation of DAG. The difference between both methylations (acidic and basic) will indicate which particular FAs were released from the *sn*-2 position and consequently those in the terminal positions. The calculation, the positional distribution determination and the data enhancement were based on a protocol described in the literature [40].

2.2.2. Sample preparation for LC–ESI-MS^2 analysis

An aliquot of cod liver oil (2 ml) was dissolved in 2 ml of chloroform:methanol (2:1, v/v), 2 ml of hexane and vortex-mixed for 30 s. The hexane phase was collected and dried under a gentle stream of nitrogen at room temperature. The dried residue was redissolved into 0.5 ml of acetonitrile:acetone (2:1, v/v). The final product was submitted to LC–ESI-MS^2 analysis. This procedure was also applied to TAG standards dissolved in chloroform:methanol (2:1, v/v).

2.3. Instrumentation

2.3.1. Gas chromatography

The GC analysis of the FAME prepared by the lipase method was performed on a Perkin-Elmer AutoSystem XL gas chromatograph (Perkin-Elmer, Norwalk, Connecticut) equipped with a liquid autosampler and a flame ionization detector. The FAME samples were analyzed on a CP-Sil 88 capillary column (50 m × 0.32 mm i.d. 0.2 μm film thickness, Varian, Courtaboeuf, France). Data collection was performed by the Perkin-Elmer TotalChrom Data System Software version 6.3. The temperature program was as follows: the oven temperature was held at 60 °C for 1 min, ramped to 160 °C at 25 °C/min, held at 160 °C for 28 min, ramped to 190 °C at 25 °C/min, held at 190 °C for 17 min, ramped to 220 °C at 25 °C/min and finally held at 220 °C for 10 min. Direct on-column injection was used. The injector port temperature was ramped instantaneously from 50 to 250 °C and the detector temperature was 250 °C. The carrier gas was ultra-pure helium at a pressure of 82 kPa. The analysis time was 60 min. This time interval was sufficient to detect FAME with chains from 10 to 24 carbons in length. The FAME peaks were identified by comparison of their retention times with the retention times of highly purified FAME standards.

2.3.2. Liquid chromatography ion-trap mass spectrometry

The LC–ESI-MS^2 used in this study was an Agilent 1100 series LC/MSD trap, SL model with an electrospray interface, a quaternary pump, degasser, autosampler, thermostatted column compartment, variable-wavelength UV detector and 10 μl injection volume. The reversed phase Ultrasphere® 5 μm Spherical 80 Å pore C-18 analytical column (250 mm × 4.6 mm i.d., Beckman Coulter, Kolbotn, Norway) was kept in the column compartment at 30 °C and the solvent system in gradient mode consisted of isopropanol: (10 mM) ammonium acetate (90:10, v/v) (A), acetone (B) and acetonitrile (C) at a flow rate of 0.8 ml/min and UV detection at 254 nm. After testing different delivered LC solvent programs, the following gradient was selected: an initial 5 min condition 90% A and 10% C that was ramped in 5 min to 65% A and 5% C and returned to the initial condition in 15 min and subsequently ramped in 5 min to 65% A and 5% C and returned to the initial condition in 30 min where it was held for 30 min.

By using this gradient program, reproducible retention times and peak areas from sample to sample were monitored. Nitrogen was used as nebulizing (50 psi) and drying gas (8 l/min) at 350 °C. The ESI source was operated in positive ion mode and the ion optics responsible for getting the ions in the ion-trap such as capillary exit, skimmer, lens and octapoles voltages were controlled by using the Smart View option with a resolution of 13000 m/z/s (FWHM/m/z = 0.6–0.7). Auto MS/MS full scan mode for 90 min in the scan range of 200–1500 m/z without dividing the acquisition program into time segments was used. The most intense ions eluting in each of the ESI-MS spectrum are automatically selected as the precursor ions for the following auto MS/MS experiments using helium as the collision gas. The product ions in ESI-MS^2 spectra are recorded and the resulting MS^2 chromatograms represent the sums of product ions from the precursor ions. Complete system control, data acquisition and processing were done using the ChemStation for LC/MSD version 4.2 from Agilent.

2.4. Computation

The identification of TAG structures in complex oils (e.g. ω-3 rich oils) is regarded as the bottleneck of LC–ESI-MS^2 analysis due to tedious and time-consuming manual calculations during the interpretation process [41,42]. To address this issue, a computational algorithm was developed to assist automatically the elucidation process.

The algorithm for the automatic interpretation of TAG molecules from LC–ESI-MS^2 data was developed by using MATLAB 7.9 [43] and the corresponding computation was performed on a Microsoft Windows XP® 2003 operating system (Microsoft Corporation, Redmond, WA, USA). The total LC + MS data (chromatograms + spectra) were exported to netCDF file and ASCII file by DataAnalysis for LC/MSD Trap Version 3.3, and were then used as the input files for the algorithm, which could automatically give the elucidation results of TAG structures without manually introducing data into the algorithm.

2.4.1. General algebraic expression for TAG elucidation

Different TAG molecules possess several common chemical groups as is shown in Scheme S1 (available in Supplementary material). For instance, (1) a common glycerol backbone (41 g/mol); (2) three methyl groups (3 × 15 g/mol); (3) three carboxylate groups (3 × 44 g/mol); (4) x, x' and x'' numbers of ethylene (–CH_2–CH_2–) groups (28 g/mol each) at *sn*-1, *sn*-2 and *sn*-3; (5) y, y' and y'' numbers of ethenyl (–CH=CH–) groups (26 g/mol each) at *sn*-1, *sn*-2 and *sn*-3 respectively. These common features are combined and used to generate a general algebraic expression for TAG elucidation.

$$[\mathrm{M}] = 41 + 3 \times 15 + 3 \times 44 + 28 \times (x + x' + x'') + 26 \times (y + y' + y'')$$

By representing the total number of ethylene and ethenyl groups as X and Y respectively,

$$X = x + x' + x'' \tag{1}$$

$$Y = y + y' + y' \tag{2}$$

it is possible to derive the general expression:

$$[\mathrm{M}] = 218 + 28 \times X + 26 \times Y \tag{3}$$

where [M] represents the TAG molecular weight (MW). It must be emphasized that X and Y should be always integral numbers (e.g. A TAG molecule containing 2.5 ethylene or 3.2 ethenyl groups does not exist). When LC–ESI-MS^2 in positive mode is used, under our experimental conditions, TAG adducts (e.g. $[M+NH_4]^+$) rather than protonated TAG molecules ($[M+H]^+$) are determined, in such a case the contribution of the ammonium (18 g/mol) should be added to Eq. (3), i.e.,

$$[\mathrm{M} + \mathrm{NH_4}]^+ = 236 + 28 \times X + 26 \times Y$$

$$X = \frac{[\mathrm{M} + \mathrm{NH_4}]^+ - 236 - 26 \times Y}{28} \tag{4}$$

By introducing the experimental m/z value of the precursor adduct ion $[M+NH_4]^+$ and substituting automatically only integral numbers of Y from 0 to 18 (the total possible range of double ethenyl bonds), it is possible to estimate X the total number of single ethylene bonds by using Eq. (4). It is important to highlight that Eq. (4) will yield a positive TAG identification if and only if Y (introduced as an integral number) is able to generate an integral X value. For example, when a TAG ammoniated adduct (m/z 890) containing three linolenic acids (18:3*n*) is analyzed, the only possible solution from Eq. (4) that yields Y and X integral values is 9 and 15 respectively (Scheme S1). Values such as 8 and 15.93 or 10 and 15.07 for Y and X are automatically rejected. The described approach is also

applicable for other types of TAG adducts. For instance, the presence of a sodiated TAG adduct $[M+Na]^+$ imply an additional contribution of the sodium (23 g/mol) to Eq. (3).

2.4.2. Computational theory for TAG interpretation

The computational theory was based on the fragmentation mechanism of TAG when using ESI-MS^2 as demonstrated in previous studies [44–46]. Briefly, the precursor adduct ions from the ESI-MS^2 mass spectrum of TAG produce very abundant DAG fragment ions due to the loss of fatty acyl moieties from the glycerol backbone. In view of the above information, the following rules were applied in the computation of TAG from the mass spectra.

1. All the observed adduct ions are of form $[M+NH_4]^+$ or $[M+Na]^+$.
2. The major product ions generated from $[M+NH_4]^+$ or $[M+Na]^+$ are DAG fragments in the form of $[M+NH_4-RCOONH_4]^+$ or $[M+Na-RCOOH]^+$ respectively, which correspond to the loss of particular FAs from the TAG backbone.
3. Only the product ions with m/z values exhibiting intensities higher than 10,000 icps (ions count per second) are screened and subjected to computation.
4. The positional distribution of the FAs on the TAG molecule is based on the relative intensities of its DAG fragments. The fatty acid which corresponds to the least abundant DAG fragment (lowest intensity) will be assigned in the *sn*-2 position on the TAG backbone. All the m/z values of possible DAG fragments observed from the mass spectrum are designated as $Frag_1$, $Frag_2$, ..., $Frag_i$, and the MW of corresponding FAs are designated as FA_1, FA_2, ..., FA_i.
5. The FA_i is calculated by subtracting $Frag_i$ from its observed precursor adduct (either $[M+NH_4]^+$ or $[M+Na]^+$) as follows:

   For $[M+NH_4]^+$ adducts:

   $$FA_i = [M+NH_4]^+ - [M+NH_4-RCOONH_4]^+ - [NH_4]^+ + [H]^+$$

   $$FA_i = [M+NH_4]^+ - Frag_i - 17 \quad (5)$$

   For $[M+Na]^+$ adducts:

   $$FA_i = [M+Na]^+ - [M+Na-RCOOH]^+$$

   $$FA_i = [M+Na]^+ - Frag_i \quad (6)$$

   The potential FAs identified by Eq. (5) or (6) are compared against their nominal MW with a tolerance of ± 0.5 m/z.
6. All the possible fatty acid candidates are combined on the TAG backbone and their theoretical X and Y values can be easily obtained by Eqs. (1) and (2) respectively. A positive TAG identification is achieved when the theoretical X and Y values are equal to those estimated from the experimental m/z value of the precursor adduct by Eq. (4).
7. The equivalent carbon number (ECN) of each identified TAG is calculated by the following equation:

   $$ECN = CN - 2Y \quad (7)$$

   where CN is the total carbon number of a TAG molecule.

In summary, the user only needs to load the exported files (netCDF file and ASCII file) into the algorithm which in turn will determine all the possible TAG molecules in the whole chromatogram fulfilling the criteria defined above.

2.5. Chemometric discrimination analysis

To examine the discrimination between genuine and adulterated cod liver oils, two different kinds of oils (marine and vegetable) were used to adulterate pure cod liver oil. The adulterants were evaluated at two different concentration levels (25 and 50%). Duplicates samples were prepared only for pure and 25% adulterated cod liver oil. The discrimination of the various samples was performed by means of principal component analysis (PCA) using their total ion current (TIC) chromatograms. The chromatogram files (1442 data points) are firstly converted into netCDF files and subsequently into Matlab files. The m/z values were rounded up to integral numbers in order to reduce the amount and complexity of the data and to allow subsequent data analysis. These chromatograms files are subjected to PCA (coded in MATLAB 7.9) after normalization. The first three scores of PCA are used to make projection plots that provide the visual discrimination between the genuine and adulterated cod liver oils.

Table 1
Positional distribution (%) of FAs on TAG from cod liver oil.

FAs	FAs composition (%)[a]			Percentage (%)[b]	
	Total%	*sn*-1 + 3%	*sn*-2%	*sn*-1 + 3%	*sn*-2%
14:0	3.93	2.71	1.22	68.89	31.11
15:0	0.42	0.35	0.07	83.55	16.45
16:0	11.88	9.29	2.59	78.17	21.83
16:1*n*-7	7.94	6.29	1.65	79.17	20.83
16:1*n*-9	0.54	0.40	0.15	73.28	26.72
16:2*n*-4	0.48	0.33	0.15	68.23	31.77
16:3*n*-3	0.30	0.12	0.18	39.61	60.39
16:4*n*-3	0.59	0.45	0.13	77.38	22.62
17:0	0.38	0.01	0.37	3.40	96.60
18:0	3.34	3.27	0.07	97.84	2.16
18:1*n*-11	1.56	1.03	0.53	66.23	33.77
18:1*n*-7	5.17	4.61	0.56	89.13	10.87
18:1*n*-9	17.56	15.17	2.39	86.39	13.61
18:2*n*-6	2.47	2.00	0.47	80.88	19.12
18:3*n*-3	0.98	0.75	0.23	76.51	23.49
18:4*n*-3	1.92	0.42	1.50	21.88	78.12
20:1*n*-11	1.35	1.08	0.27	80.04	19.96
20:1*n*-7	0.42	0.33	0.09	78.76	21.24
20:1*n*-9	9.95	7.66	2.29	77.00	23.00
20:2*n*-6	0.31	0.22	0.09	71.40	28.60
20:4*n*-3	0.68	0.24	0.45	34.58	65.42
20:4*n*-6	0.54	0.16	0.38	29.64	70.36
EPA	8.54	2.11	6.43	24.72	75.28
22:1*n*-11	6.23	4.66	1.58	74.72	25.28
22:1*n*-9	0.89	0.80	0.10	89.16	10.84
DPA	1.30	0.31	0.99	24.19	75.81
DHA	9.55	0.40	9.18	4.07	96.04
24:0	0.18	0.11	0.07	60.37	39.63
24:1*n*-9	0.58	0.25	0.34	42.15	57.85

[a] Each value represents the mean value of duplicates (Total: total FAs on all the positions; *sn*-2%: FAs on *sn*-2 position; *sn*-1 + 3%: FAs on both *sn*-1 and *sn*-3 positions).
[b] *sn*-1 + 3% = (*sn*-1 + 3/Total) × 100%, *sn*-2% = (*sn*-2/Total) × 100%.

3. Results and discussion

3.1. Lipase stereospecific analysis

The positional distribution of FAs in the TAG of cod liver oil obtained by the benchmark lipase method is shown in Table 1. The total FAs composition analysis indicated that cod liver oil is principally characterized by 18:1*n*-9 (17.56%), 16:0 (11.88%), 20:1*n*-9 (9.95%), DHA (9.55%) and EPA (8.54%). In addition, the results in Table 1 showed that ω-3 FAs such as DHA (96.04%), 18:4*n*-3 (78.12%), DPA (75.81%), EPA (75.28%), 20:4*n*-3 (65.42%) and 16:3*n*-3 (60.39%) are mainly located at the *sn*-2 position of TAG species. A published stereospecific analysis of cod liver oil of the same brand used in the present article and by ^{13}C NMR [13] failed to detect 20:4*n*-3 and DPA. In addition, this reported study found that EPA and 18:4*n*-3 were equally distributed on the three stereospecific positions of TAG species. The only result in agreement with the present lipase method (Table 1) was DHA primarily at the *sn*-2 position.

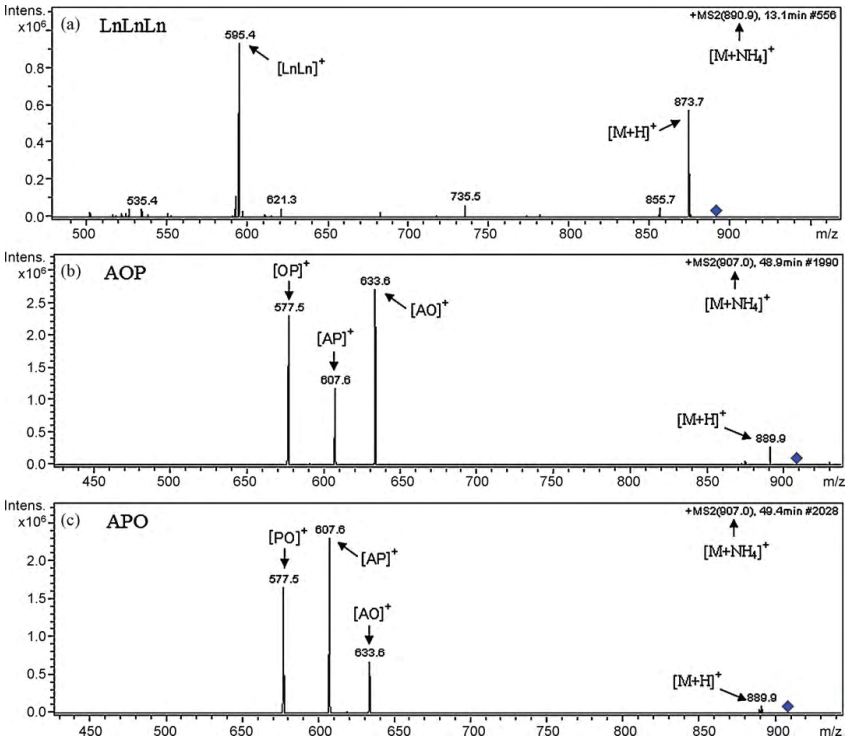


Fig. 1. ESI-MS^2 spectra of the ammoniated TAG standards: (a) LnLnLn, (b) AOP and (c) APO.

3.2. Elucidation of TAG in standards and vegetable oils by LC–ESI-MS^2

The performance of the developed TAG elucidation algorithm was firstly tested by using TAG standards. It is important to mention that the preferential cleavage fragmentation mechanisms by ESI-MS^2 to be discussed below have been demonstrated previously [44–46] and incorporated in the algorithm. The following examples will illustrate the interpretation function as well as the behaviour of TAG mass spectra.

A TAG molecule with the same fatty acid on its backbone, such as LnLnLn, exhibits a very simple mass spectrum (Fig. 1a) with only a single DAG fragment ion ([LnLn]$^+$ at m/z 595.4) resulting from the dissociation of linolenic acid (18:3n, Ln) from the LnLnLn. A different pattern arises from a TAG molecule containing three different acyl groups such as AOP. The AOP ammoniated precursor $[M+NH_4]^+$ at m/z 907 (Fig. 1b) gives rise to three DAG fragments $[OP]^+$, $[AP]^+$ and $[AO]^+$ at m/z 577.5, 607.6 and 633.6 respectively. The least abundant DAG fragment ion, at m/z 607.6, corresponds to the loss of oleic acid (18:1n, O) from the middle position (*sn*-2), indicating that the cleavage from this particular position is energetically less favoured than the outer positions (*sn*-1 and *sn*-3). Similarly, the mass spectrum of APO (Fig. 1c) displays the same three DAG fragment ions observed in the mass spectrum of its stereoisomer AOP, however the relative intensities of the generated DAG fragments are different in both spectra. In the case of APO (Fig. 1c), the DAG fragment $[AO]^+$ at m/z 633.6 displays the lowest intensity, indicating the loss of palmitic acid (16:0, P) from the *sn*-2 position. The observed ESI-MS^2 preferential cleavage of the FAs from the outer positions and the relative low intensity at the middle position of the DAG fragments which enables assigning a particular fatty acid to the *sn*-2 position have been generally investigated by means of TAG standards [44–46].

The elucidation capability of the proposed algorithm was also tested by using commercial linseed and rapeseed oils. It must be said that published reports on the elucidation of TAG species of these particular oils by LC atmospheric pressure chemical ionization single MS (LC–APCI-MS) are generally based on the above described preferential cleavage [47,48]. The elucidated TAG structures by using the developed algorithm for linseed and rapeseed oils were in accordance with those reported elsewhere [47,49–51]. The positional distribution of FAs in TAG and the elucidated TAG species of these vegetable oils are listed in the Supplementary material.

3.3. Elucidation of TAG in cod liver oil by LC–ESI-MS^2

The TAG species in the cod liver oil are identified by exporting simultaneously the total LC + MS data (chromatograms + spectra) into the developed algorithm where the mass spectra are elucidated and associated automatically to specific retention times.

The TIC chromatogram of cod liver oil and associated ECN values is shown in Fig. 2. The various elucidated TAG structures described in Table 2 are listed in increasing order of ECN along with their *sn*-2 and *sn*-1/3 positions (no distinction is made between the outer positions). Table 2 revealed that the FAs exhibiting the highest relative concentrations in Table 1 (lipase method) namely, 16:0, 16:1n, 18:1n, 20:1n, 22:1n, EPA and DHA were the most frequent detected in the various TAG structures.

Several examples for the identification of TAG species in cod liver oil are given to illustrate the interpretation process of the algorithm.

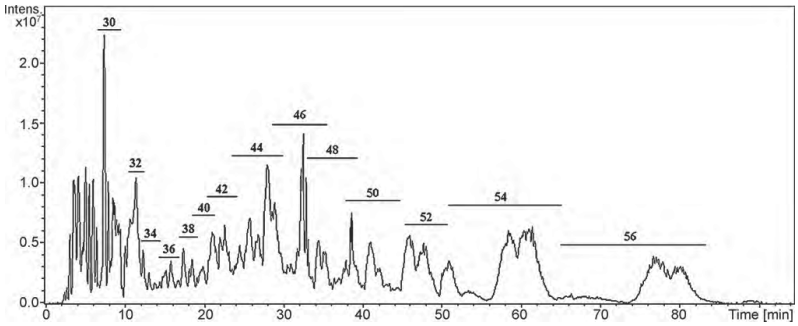


Fig. 2. TIC chromatogram of cod liver oil with the associated ECN values.

3.3.1. Elucidation of single TAG structures in cod liver oil

The ESI-MS^2 spectrum of an ammoniated TAG adduct obtained at 16.3 min is shown in Fig. 3a. The precursor ion $[M+NH_4]^+$ at *m*/*z* 968.9 produces six possible DAG fragments which can be easily visualized in the mass spectrum. The algorithm firstly, arranges the potential DAG fragments in descending order of intensity, namely *m*/*z* 649.5, 623.5, 631.4, 621.5, 669.4, 606.9 (Fig. 3b) and after performing the various computation rules previously described it indicates that four out of six fragments, specifically *m*/*z* 649.5, 623.4, 669.4 and 621.5 result from the loss of EPA, DHA, 18:1*n* and DPA from potential TAG ammoniated precursors respectively, while the masses at *m*/*z* 320.49 and 344.99 estimated from the fragments at *m*/*z* 631.4 and 606.9 respectively do not match any saturated or unsaturated FAs containing between 14 and 35 carbon molecules. The algorithm identified the combination EPA, DHA and 18:1*n* as a TAG molecule. This combination fulfils all the requirements described in Section 2.4. In addition, the algorithm assigned the *sn*-2 position to 18:1*n* as a result of the low intensity of the corresponding fragment at *m*/*z* 669.4. Although fragment D (*m*/*z* 621.5) (Fig. 3a) seems to correspond with the loss of DPA, this particular fatty acid does not comply with the general requirements for a positive TAG identification described in Section 2.4. The calculation of the total number of ethylene (*X*) and ethenyl (*Y*) group excludes automatically DPA from the precursor ion $[M+NH_4]^+$ at *m*/*z* 968.9. All the combinations containing DPA cannot yield the integral numbers 15 and 12 for *X* and *Y* respectively. The presence of the fragment at *m*/*z* 621.5 might be due to the interference from other TAG fractions.

3.3.2. Elucidation of TAG positional and structural isomers in cod liver oil

The analysis of complex mixtures, such as cod liver oil, by LC–ESI-MS^2 brings about the presence of overlapping chromatographic peaks corresponding to positional or structural isomers. For instance, the extracted ion chromatogram (EIC) of the sodiated precursor ion at *m*/*z* 927.9 (Fig. 4) exhibits two chromatographic peaks overlapping at 22.6 and 22.8 min. Although the mass spectra of these peaks display similar fragmentation patterns at *m*/*z* 577.5, 599.5, 623.4, 645.4, 671.5 and 699.5, their relative intensities are different, indicating the presence of stereoisomers. The algorithm revealed that only the combination of 16:0, 18:1*n* and DHA constitutes a positive TAG molecule in both spectra (Fig. 4a and b) and that 16:0 and DHA (the least intense fragments) are located in the *sn*-2 position of the identified TAG positional isomers at 22.6 and 22.8 min respectively. It is important to mention that the sodiated adducts observed in Fig. 4 might be ascribed to some sodium impurities in the solvents which have been reported elsewhere [52–54].

The LC–ESI-MS^2 analysis of cod liver oil also revealed the presence of structural isomers. For instance, although the EIC at *m*/*z* 877.0 exhibits one chromatographic peak at 32.8 min (Fig. 5a), the algorithm shows firstly, that the four DAG fragment ions (*m*/*z* 577.5, 603.5, 605.6 and 549.5) derived from the precursor ion $[M+NH_4]^+$ at *m*/*z* 877 (Fig. 5a) result from the loss of 18:1*n*, 16:0, 16:1*n* and 20:1*n* from TAG molecules and secondly that with these identified FAs only two TAG species fulfil the algorithm criteria, namely 18:1*n*/16:0/18:1*n* and 16:0/20:1*n*/16:1*n* (*sn*-2 positions are underlined). Similarly, the ability of the algorithm to identify co-eluting sodiated TAG isomers from a single chromatographic peak is showed in Fig. 5b where the two TAG molecules fulfilling the algorithm criteria are 18:1*n*/DHA/20:1*n* and 16:1*n*/22:1*n*/DHA.

3.4. Comparison with other LC–ESI-MS^2 studies

Although plant oils are the most studied samples by LC–ESI-MS^2, little information is given regarding how the reported TAG species were identified [35–39,55]. For instance, Svensson and Adlercreutz [55] identified 12 TAG species in the transesterified blend of rapeseed and butter oils, however, the identification of TAG was not explained. Complex samples have been also studied by LC–ESI-MS^2 [41,42]. For instance, Kalo et al. [41] reported the determination of TAG in butterfat by normal-phase LC–ESI-MS^2, where they analyzed four fractions of butterfat separated by solid phase extraction and subsequently identified 450 TAG species in total. However, the details regarding the identification of TAG species were not sufficiently illustrated. Our investigation explains the derivation of the rules for TAG elucidation by LC–ESI-MS^2 in conjunction with the proposed algorithm, based on TAG structural features and fragmentation mechanisms. Typical examples for the elucidation of positional and structural isomers of TAG structures are also provided, which gives a full overview of the interpretation of intact TAG molecules determined by LC–ESI-MS^2.

3.5. Chemometric detection of adulteration

The converted data points of the TIC chromatograms were studied by PCA to evaluate if the TAG information contained in the TIC chromatograms enables the discrimination of pure from adulterated cod liver oil. The 3D score plot (Fig. 6) explains 75.4% of the total data variation and provides a clear differentiation between genuine and adulterated cod liver oils. The pure cod liver oil samples (designated as CLO) are clustered together and clearly separated from cod liver oil adulterated with soy oil (CLO/SOY) or seal oil (CLO/SEAL) at the two levels of impurities added in this study (25 and 50%). In general, the CLO/SEAL samples in Fig. 6 are closer to pure CLO samples compared to CLO/SOY. This behaviour could

Table 2
TAG species identified by LC–ESI-MS^2 in cod liver oil. Note that no distinction is made between *sn*-1 and *sn*-3 positions.

ECN	Identified TAG species							
30	EEE*	StDE*	EDE*	DStD				
32	ELnE*							
34	MDSt*	PEHt	PoRE	PoStSt*	PoStD	PoESt*	PoDR	PoDE*
	HtPD	HtDO	OHtE	LDSt	LDE*	LnLnSt	LnLnD	StME
	StPoE	StPoD	StLSt	StLE	ArLnE	ArArE	EME*	EMD
	EPoE*	EPoD	ELE*	ELD	DPoD			
36	MLnD	MStDo	MArE	MArD	MDoSt	MDoE	PStE	PStD
	PEE*	PDR	PDSt*	PDE*	PoLnE	PoLnD	PoStAr	PoArE
	PoArD	PoDPR	PoDoSt	PoDLn*	ROE	RDO	HtArO	ORD
	ORE	OStSt	OStE*	OStD	OESt	ODSt	ODE*	LArSt
	LnPoD	LnHDo	LnLE	LnLnLn	LnLnAr	LnArLn	LnArAr	StPD
	StPE	StPoDo	StOSt	StOD	StOE	StLAr	ArPoE	ArArAr
	EMDo	EPE*	EPD	EOE*	EOD	EEcE	DPD	DOD
38	MLE	MLD	MLnLn	MEPo*	MDPo*	PRLn	PArD	PDoE
	PDAr	PoME*	PoMD	PoPoE*	PoLE	PoLnLn	PoStPo	PoEPo*
	PoDPo*	RLnO	HtOL	SStD	OLnD	OArE	ODLn	LLSt
	LLD	GRD	ArPD	ArOE	ArLAr	EPDo	ESE	
40	MOE*	MOD	MArPo	MArL	MEO*	MDoM	MDoPo	MDoL
	MDP*	MDO*	PME*	PPoSt	PPoE*	PPoD	PLE	PLD
	PLnLn	PStPo	PStL	PArAr	PDPo*	PDL*	PoPO*	PoPE*
	PoPD	PoRO	PoOE*	PoOD	PoStO	PoArPo	PoArL	PoEO*
	PoDoPo	PoDoL	PoDO*	HSSt	HGD	RAE	OMSt	OME*
	OMD	OPoE*	OPoD	OHDo	OHtO	OLSt	OLE*	OLnLn
	OStL	OEL*	ODL*	LPE*	LPD	LOSt	LLLn	LLnL
	LArL	LEG*	LDoL	AHtAr	GPoD			
42	MSE	MSD	MGSt	MGE*	MGD	MArO	MES	MDoP
	MDoO	PtPtDo	PMDo	PPoAr	POSt	POE	PLAr	PStO
	PGHt	PEcE	PArPo	PEP*	PEO*	PDoPo	PDoL	PDP*
	PDO*	PoPDo	PoSSt	PoSD	PoLL	PoStG	PoGSt	PoGE*
	PoGD	PoEcAr	PoArO	PoDoO	HHG	HArG	SME*	SMD
	SPoSt	SPoD	SLnLn	SArLn	OMAr	OMDo	OPSt	OPE*
	OPD	OPoLn	OHtG	OStO	OArL	OEO*	ODO*	LLL
	LnGLn	StMG	StPoG	GME*	GMD	GPoD		
44	MAD	MGAr	MErSt	MErD	MDoS	MDoG	PSD	POAr
	PStG	PGSt	PGE*	PGD	PArO	PES*	PEG*	PDoP
	PDpO	PDS*	PDG*	PoSDo	PoOPo*	PoStEr	PoGAr	PoArG
	PoErSt	PoErD	PoDoS	MaMaD	SPD	SPoAr	SPoDo	SHtG
	SOSt	SOE*	SOD	SStO	SEO*	SDO*	OPAr	OPDo
	ORG	OHtEr	OSSt	OSE	OLnO	OStG	OGSt	OGD
	OArO	OEG*	ODS*	ODG*	LnLnEr	LnALn	StPoEr	AMD
	GMAr	GMDo	GPE*	GPD	GPoAr	GHtG	GOD	EMEr
	ErMD	ErPoD						
46	MHEr	MGPo*	MGL*	MEcO	MBD	MErH	MErDo	MNE
	MDPEr	PPoO*	PHG	PSLn	POL*	PLO	PLnS	PAD
	PGH	PGAr	PGDo	PEcPo	PArG	PEEr	PErSt	PErE
	PErD	PDoS	PDoG	PDEr	PoPO*	PoPoG*	PoSPo	PoSL
	PoADo	PoGPo*	PoArEr	PoErAr	PoND	HSO	SPoL*	SHO
	SODo	SGSt	SGE*	SGD	SEcAr	SEG*	SDoO	SDS*
	OMO*	OPL*	OPoO*	OSDo	OStEr	OAE*	OArG	OErSt
	OErD	ODEr	LMG*	LnNLn	StSG	StGG	APD	AOE*
	GPDo	GHtEr	GSE	GSD	GStG	GGD	GEG*	GDG*
	EPEr	BMD	ErMDo	ErPD				
48	MErPo	PPoG*	PHEr	PEcO	PDoEr	PNE	PND	PoMEr
	PoAL*	PoGO*	PoEcS	PoErPo	SHG	SOL*	SLO	SLnS
	SEEr	SErD	SDoG	SDEr	OMG*	OPO*	OPoG*	OSL
	OOO*	OArEr	ONSt	OND	LMEr	LPG*	StErG	AGD
	GMEc	GHtN	GSDo	GStEr	GAD	GArG	GErD	GDEr
	ArPEr	ArOEr	ErPDo	ErHtEr	ErSD	ErGD	DPMN	DPN
	DON							
50	MEcEr	MErO	PMEr	PPoEr	PSO	POS*	PLEr	PGP*
	PGO*	PGEc	PErPo	PErL	PDoN	PNDo	PoMG*	PoSG
	PoOEr	PoAEc	PoGG*	PoErO	PoNPo	SMG*	SPoG*	SHEr
	SOEc	SGL*	SDPEr	OMEr	OPG*	OPoEr	OSO	OOG*
	OAL*	OGO*	LMN	LPEr	LSG	LOA*	StGN	StNG
	AAD	GMG*	GPoG*	GArEr	GND	ArON	ArGEr	ErStEr
	ErErD	ErDEr	DPPN	DGN				
52	MAG*	MErG	MNO	PMN	PPoN	POEr	PAO*	PGS*
	PGG*	PErP	PErO	PErG	PoLiPo	SMEr	SOS*	OPEr
	OPoN	OSG	OGG*	OErO	GMEr	GPG*	GPoEr	GOG*
	ErDN	ErND						
54	MAEr	MGB*	POB*	PON	PGA*	PErS	PErG	PErO
	PNO	SPEr	SOEr	SAO*	SGS*	OSEr	OGEr	OErG
	ONO	GMN	GPEr	GSG	GOEr	GGG*	ErMEr	ErPoEr
56	MNEr	PGN	PNG	PoNEr	ONG	ONEr	GSEr	GGEr
	GErG	GNG	ErPEr	ErPoN	ErOEr	ErGS	ErGEr	ErON

Note: *major TAG species.
Abbreviations: M: 14:0; Pt: 15:0; P: 16:0; Po: 16:1*n*; H: 16:2*n*; R: 16:3*n*; Ht: 16:4*n*; Ma: 17:0; S: 18:0; O: 18:1*n*; L: 18:2*n*; Ln: 18:3*n*; St: 18:4*n*; A: 20:0; G: 20:1*n*; Ec: 20:2*n*; Ar: 20:4*n*; E: EPA; B: 22:0; Er: 22:1*n*; DPA: Do; DHA: D; Li: 24:0; N: 24:1*n*.

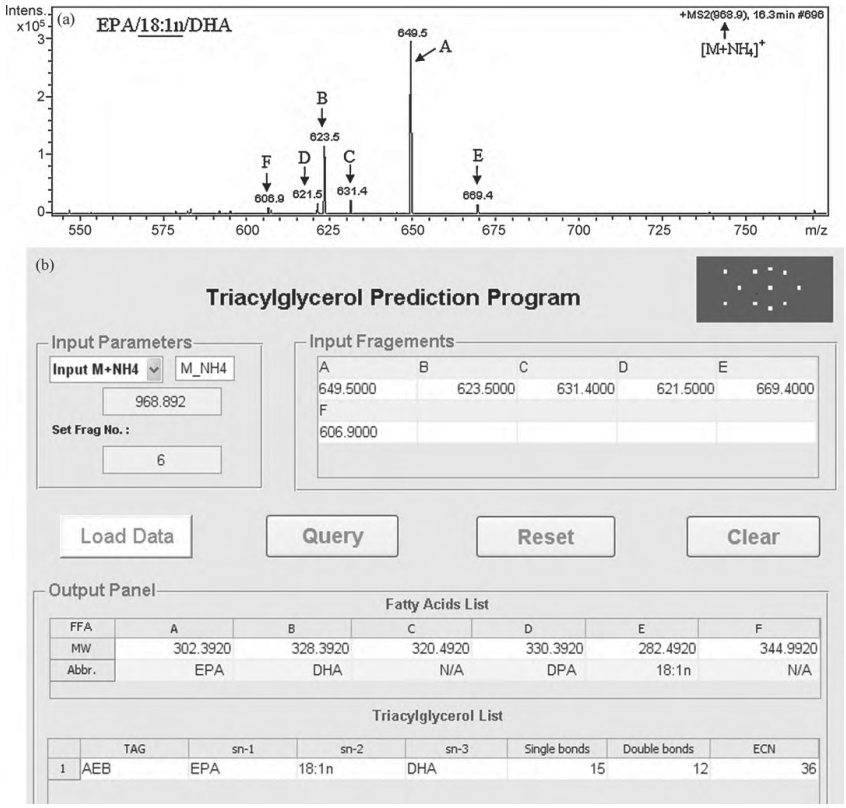


Fig. 3. (a) ESI-MS^2 spectrum of the ammoniated EPA/18:1*n*/DHA (*m*/*z* 968.9) obtained at 16.3 min of cod liver oil. (b) Algorithm outcomes of the above data at 16.3 min.

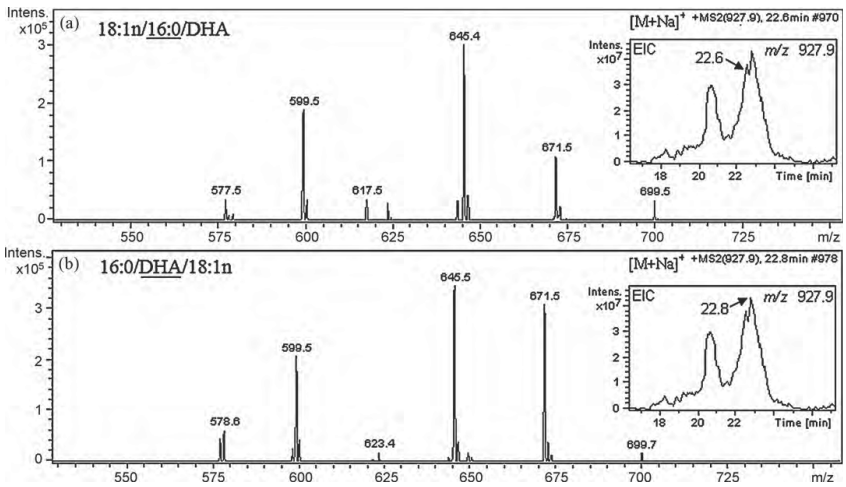


Fig. 4. ESI-MS^2 spectra of the sodiated adducts from cod liver oil: (a) 18:1*n*/16:0/DHA at 22.6 min and (b) 16:0/DHA/18:1*n* at 22.8 min and their corresponding embedded EIC at *m*/*z* 927.9.

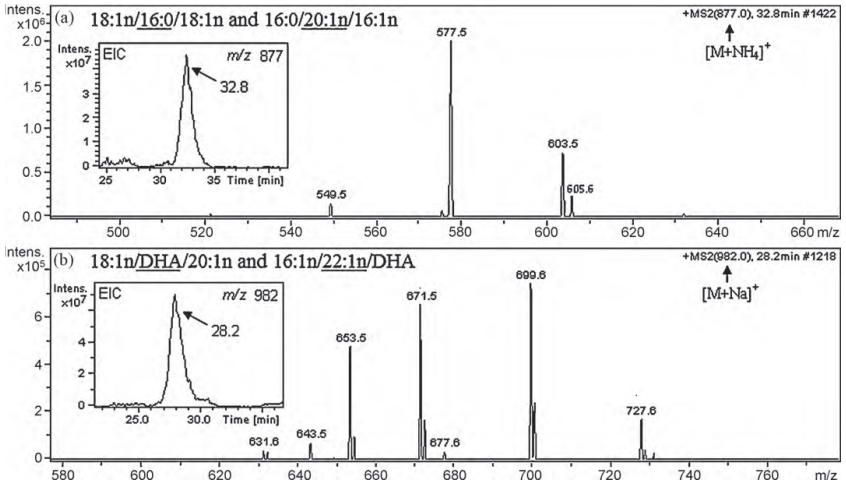


Fig. 5. (a) ESI-MS^2 spectrum of the ammoniated adducts from cod liver oil 18:1*n*/16:0/18:1*n* and 16:0/20:1*n*/16:1*n* at 32.8 min and their corresponding embedded EIC at *m*/*z* 877.0; (b) ESI-MS^2 spectrum of the sodiated adducts from cod liver oil 18:1*n*/DHA/20:1*n* and 16:1*n*/22:1*n*/DHA at 28.2 min and their corresponding embedded EIC at *m*/*z* 982.0.

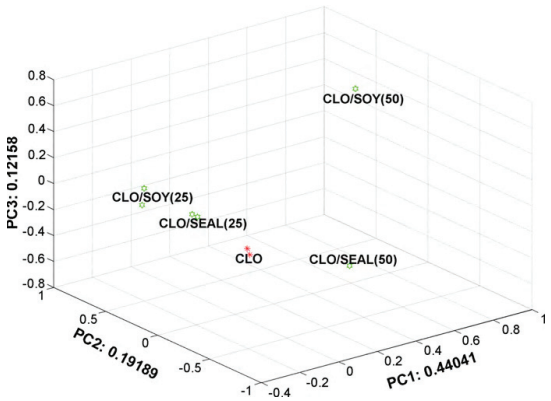


Fig. 6. PCA score plot of genuine and adulterated cod liver oil based on the LC–ESI-MS^2 analysis. (CLO: cod liver oil; SOY: soy oil; SEAL: seal oil. The numbers in bracket represent the concentrations of adulterant in cod liver oil.)

be ascribed to the lack of ω-3 polyunsaturated FAs (PUFAs) in soy oil. The detection of seal oil as adulterant of cod liver oil is regarded as exceedingly difficult due to their strong resemblance [56–59]. However, the developed algorithm, for elucidating TAG structures, revealed ω-3 PUFAs mainly located at the *sn*-2 position in pure cod liver oil, while for CLO/SEAL (25 or 50) the algorithm revealed ω-3 PUFAs not only at the *sn*-2 positions but also at the *sn*-1/3 positions which clearly indicated the presence of seal oil. It has been reported that ω-3 PUFAs are preferentially located at the terminal positions of TAG in seal oil [7,10]. The differences in TAG structures from CLO and CLO/SEAL samples elucidated by the algorithm were substantiated by the PCA discrimination study (Fig. 6).

4. Conclusion

A LC–ESI-MS^2 strategy was successfully established to directly identify the relative arrangement of the acyl groups on the glycerol backbone of cod liver oil. The developed computational algorithm facilitated the rapid structural elucidation of the TAG molecules in cod liver oil based on the information obtained from the LC–ESI-MS^2 data. The combined information from the lipase and LC–ESI-MS^2 approach enable a full examination not only on the total FAs composition but also on the specific positioning of FAs on intact TAG molecules in cod liver oil which represents a useful means to help the understanding of its properties and nutritional value as well as the detection of adulteration for these kinds of products.

Acknowledgements

The European Commission in the context of the Erasmus Mundus Program and The Norwegian Research Council (SIP project NRF 173534/I30) are gratefully acknowledged for financial support of Y.Z. and Z.D. respectively. The authors would like to thank Yan-Chun Ho for technical assistance in programming and Tormod Bjørkkjær for kindly donating the soy and seal oil samples.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.talanta.2010.06.055.

References

[1] H. Shimokawa, P.M. Vanhoutte, Circulation 78 (1988) 1421–1430.
[2] B. Osterud, E. Elvevoll, H. Barstad, J. Brox, H. Halvorsen, K. Lia, J.O. Olsen, R.L. Olsen, C. Sissener, O. Rekdal, E. Vognild, Lipids 30 (1995) 1111–1118.
[3] J. Gruenwald, H.J. Graubaum, A. Harde, Adv. Ther. 19 (2002) 101–107.
[4] L.A. Brunborg, T.M. Madland, R.A. Lind, G. Arslan, A. Berstad, L. Frøyland, Clin. Nutr. 27 (2008) 614–622.
[5] B. Galarraga, M. Ho, H.M. Youssef, A. Hill, H. McMahon, C. Hall, S. Ogston, G. Nuki, J.J.F. Belch, Rheumatology 47 (2008) 665–669.
[6] H. Brockerhoff, R.J. Hoyle, P.C. Hwang, C. Litchfield, Lipids 3 (1968) 24–29.
[7] Y. Ando, T. Ota, Y. Matsuhira, K. Yazawa, J. Am. Oil Chem. Soc. 73 (1996) 483–487.
[8] J.J. Myher, A. Kuksis, K. Geher, P.W. Park, D.A. DiersenSchade, Lipids 31 (1996) 207–215.
[9] C.V. Nwosu, L.C. Boyd, J. Food Lipids 4 (1997) 65–74.
[10] U.N. Wanasundara, F. Shahidi, J. Food Lipids 4 (1997) 51–64.
[11] Y. Ando, M. Satake, Y. Takahashi, Lipids 35 (2000) 579–582.
[12] F.D. Gunstone, S. Seth, Chem. Phys. Lipids 72 (1994) 119–126.
[13] M. Aursand, L. Jørgensen, H. Grasdalen, J. Am. Oil. Chem. Soc. 72 (1995) 293–297.

[14] M.D. Guillén, I. Carton, E. Goicoechea, P.S. Uriarte, J. Agric. Food Chem. 56 (2008) 9072–9079.
[15] F. Turon, F. Bonnot, Y.C.M. Pina, J. Graille, Chem. Phys. Lipids 125 (2003) 41–48.
[16] F. Turon, Oleag. Corps Gras Lipides 10 (2003) 144–149.
[17] H. Brockerhoff, Lipids 6 (1971) 942–956.
[18] M. Yurkowski, H. Brockerhoff, Biochim. Biophys. Acta 125 (1966) 55–59.
[19] G. Andreotti, E. Trivellone, R. Lamanna, A.D. Luccia, A. Motta, J. Dairy Sci. 83 (2000) 2432–2437.
[20] M.R.V. Calsteren, C. Barr, P. Angers, J. Arul, Bull. Magn. Reson. 18 (1996) 175–177.
[21] S. Kubow, J. Nutr. Biochem. 7 (1996) 530–541.
[22] W. Chakra, Food Sci. Technol. 20 (2008) 199–202.
[23] G. Nelson, R. Ackman, Lipids 23 (1988) 1005–1014.
[24] W. Neff, E. Selke, T. Mounts, W. Rinsch, E. Frankel, M. Zeitoun, J. Am. Oil Chem. Soc. 69 (1992) 111–118.
[25] E. Weise, USA Today, Retrieved from, 23 January 2009 (15.03.10) http://www.usatoday.com/news/health/2009-01-19-fake-foods_N.htm.
[26] Z. Hossain, The Daily Independent Bangladesh (Internet Edition), Retrieved from, 13 March 2005 (15.03.10) http://www.theindependent-bd.com/details.php?nid=162455.
[27] O.C.S. Carter, Am. Philos. Soc. 22 (1885) 296–299.
[28] E.J. Parry, Lancet 163 (1904) 378.
[29] M. Lees, Food Authenticity and Traceability, first ed., Woodhead Publishing Ltd., Cambridge, 2003.
[30] I. Martinez, D. James, H. Loréal, Application of Modern Analytical Techniques to Ensure Seafood Safety and Authenticity, FAO Fisheries Technical Paper 455, Rome, 2005.
[31] J.S. Perona, V. Ruiz-Gutiérrez, J. Liq. Chromatogr. Relat. Technol. 22 (1999) 1699–1714.
[32] M. Hori, Y. Sahashi, S. Koike, R. Yamaoka, M. Sato, Anal. Sci. 10 (1994) 719–724.
[33] A.S. Mcgill, C.F. Moffat, Lipids 27 (1992) 360–370.
[34] K. Takahashi, T. Hirano, M. Saito, Nippon Suisan Gakk 54 (1988) 523–528.
[35] S.D. Segall, W.E. Artz, D.S. Raslan, V. Ferraz, J.A. Takahashi, J. Am. Oil Chem. Soc. 81 (2004) 143–149.
[36] S.D. Segall, W.E. Artz, D.S. Raslan, G.N. Jham, J.A. Takahashi, J. Agric. Food Chem. 53 (2005) 9650–9655.
[37] S.D. Segall, W.E. Artz, D.S. Raslan, P.F. Vany, A.T. Jacqueline, J. Sci. Food Agric. 86 (2006) 445–452.
[38] J.L. Gómez-Ariza, A. Arias-Borrego, T. García-Barrera, R. Beltran, Talanta 70 (2006) 859–869.
[39] L. Heidi, S. Jukka-Pekka, K. Heikki, Rapid Commun. Mass Spectrom. 21 (2007) 2361–2373.
[40] J.P. Williams, M.U. Khan, D. Wong, J. Lipid Res. 36 (1995) 1407–1412.
[41] P. Kalo, A. Kemppinen, V. Ollilainen, Lipids 44 (2009) 169–195.
[42] K. Ikeda, Y. Oike, T. Shimizu, R. Taguchi, J. Chromatogr. B 877 (2009) 2639–2647.
[43] MATLAB 7.9 (R2009b), The Math Works Inc., 2009.
[44] C. Cheng, M.L. Gross, E. Pittenauer, Anal. Chem. 70 (1998) 4417–4426.
[45] E. Hvattum, Rapid Commun. Mass Spectrom. 15 (2001) 187–190.
[46] L. Marzilli, L. Fay, F. Dionisi, P. Vouros, J. Am. Oil Chem. Soc. 80 (2003) 195–202.
[47] M. Holčapek, P. Jandera, P. Zderadicka, L. Hrub, J. Chromatogr. A 1010 (2003) 195–215.
[48] H.R. Mottram, S.E. Woodbury, R.P. Evershed, Rapid Commun. Mass Spectrom. 11 (1997) 1240–1252.
[49] J.D.J. van den Berg, N.D. Vermist, L. Carlyle, M. Holcapek, J.J. Boon, J. Sep. Sci. 27 (2004) 181–199.
[50] R.M. Hazel, E.W. Simon, P.E. Richard, Rapid Commun. Mass Spectrom. 11 (1997) 1240–1252.
[51] M. Lísa, M. Holčapek, J. Chromatogr. A 1198 (2008) 115–130.
[52] K.L. Duffin, J.D. Henion, J.J. Shieh, Anal. Chem. 63 (1991) 1781–1788.
[53] K. Hartvigsen, A. Ravandi, K. Bukhave, G. Holmer, A. Kuksis, J. Mass Spectrom. 36 (2001) 1116–1124.
[54] S.D. Segall, W.E. Artz, D.S. Raslan, V.P. Ferraz, J.A. Takahashi, Food Res. Int. 38 (2005) 167–174.
[55] J. Svensson, P. Adlercreutz, Eur. J. Lipid Sci. Technol. 110 (2008) 1007–1013.
[56] H.G. Greenish, A Text Book of Materia Medica, Being an Account of the More Important Crude Drugs of Vegetable and Animal Origin, J. & A. Churchill, 1920.
[57] R. Benedikt, Chemical Analysis of Oils, Fats, Waxes, and of the Commercial Products Derived Therefrom, General Books, 2010.
[58] H. Richardson Procter, The Principles of Leather Manufacture, Nabu Press, 2010.
[59] J.C. Drummonda, S.S. Zilvaa, J. Goldinga, J. Agric. Sci. 13 (1923) 153–162.

Paper III

Least squares spectral resolution of liquid chromatography–mass spectrometry data of glycerophospholipids

Y.X. Zeng, S.A. Mjøs, S. Meier, C.C. Lin, R. Vadla
Journal of Chromatography A 1280: 23–34 (2013).

III

Journal of Chromatography A, 1280 (2013) 23–34

Contents lists available at SciVerse ScienceDirect

Journal of Chromatography A

journal homepage: www.elsevier.com/locate/chroma

Least squares spectral resolution of liquid chromatography–mass spectrometry data of glycerophospholipids

Ying-Xu Zeng[a], Svein A. Mjøs[a,*], Sonnich Meier[b], Chen-Chen Lin[a], Reidun Vadla[a,b]

[a] Department of Chemistry, University of Bergen, P.O. Box 7803, N-5020 Bergen, Norway
[b] Institute of Marine Research, P.O. Box 1870 Nordnes, NO-5817 Bergen, Norway

ARTICLE INFO

Article history:
Received 4 July 2012
Received in revised form
20 December 2012
Accepted 26 December 2012
Available online 9 January 2013

Keywords:
Glycerophospholipids
Phosphatidylcholine
Phosphatidylethanolamine
Sphingomyelins
LC–MS
Least squares spectral resolution

ABSTRACT

Liquid chromatography–mass spectrometry represents a powerful tool for the analysis of intact glycerophospholipids (GPLs), but manual data interpretation may be a bottleneck in these analyses. The present paper proposes a least square regression approach for the automated characterization and deconvolution of the main GPLs species, *i.e.*, phosphatidylcholine and phosphatidylethanolamine analyzed by class-specific scanning methods such as precursor ion scanning and neutral loss scanning, respectively. The algorithm is based on least squares resolution of spectra and chromatograms from theoretically calculated mass spectra, and eliminates the need for isotope correction. Results from the application of the methodology on reference compounds and extracts of cod brain and mouse brain are presented.

1. Introduction

Glycerophospholipids (GPLs) are the building blocks of cell membranes that are present in all organisms. They are dynamically involved in many important biological functions and processes such as membrane trafficking, cellular signaling and metabolic regulation [1]. An increasing body of evidence suggests that the chemical diversity and composition of GPLs are correlated to certain diseases, such as Alzheimer's disease [2] and other neural disorders [3], cardiovascular disease, and immunological abnormalities [4].

GPLs can be classified into several classes according to the polar head group linked to the phosphate group at the *sn*-3 position of the glycerol backbone. Among these classes phosphatidylcholine (PC) and phosphatidylethanolamine (PE) are the dominant species in most eukaryotic membranes, present in about 3:2 molar ratio and constitute around 75 mol% of total GPLs [5]. They can be further defined by different combinations of fatty acids with various carbon numbers and double bounds, esterified at the *sn*-1 and *sn*-2 positions of the backbone, thus resulting in a large number of PC and PE molecular species. Since the roles of GPLs in cellular biochemistry are still only partly understood, great efforts have been initiated with focus on the systematic analyses of GPLs and with emphasis on the major PC and PE species in order to provide insights into the biological and physiological functions of the GPLs in biological systems.

Many indirect and direct approaches have been used for the analysis of PC and PE molecular species. The former normally consist of separation of PC and PE by thin layer chromatography (TLC), liquid chromatography (LC) or solid phase extraction (SPE), followed by derivatization of individual fatty acids in the different lipid classes and quantification by gas chromatography (GC). The main disadvantages of these indirect approaches not only lie in multiple laborious steps involved, poor resolution and reproducibility along with possible oxidation of the fatty acids, but also in the inability of providing structural information of intact PC and PE species [6]. On the other hand, direct approaches utilizing high performance liquid chromatography (HPLC) coupled to various detectors such as ultraviolet (UV), evaporative light-scattering detectors (ELSD) and mass spectrometry (MS), are today widely employed due to their capability of analyzing the intact PC and PE species [7–9]. Advances in soft ionization techniques such as electrospray ionization (ESI) have led to rapid progress in MS-based approaches including the direct infusion technique (shotgun lipidomics) and liquid chromatography coupled to MS (LC–MS), making them popular lipidomic platforms [10]. In particular, the GPLs class-specific MS profiling method is advantageous because it allows the specific detection of various GPLs classes based on their fragmentation mechanisms with high selectivity and

* Corresponding author. Tel.: +47 5558 3553; fax: +47 5558 9490.
E-mail address: svein.mjos@kj.uib.no (S.A. Mjøs).

0021-9673/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.chroma.2012.12.070

sensitivity. For example, by using precursor ion scanning (PIS) of m/z 184 with positive ionization on a triple quadrupole instrument, all the PC species can be detected since the phosphocholine head group is normally lost as a charged fragment, while by using neutral loss scanning (NLS) of m/z 141 with positive ionization, all the PE species can be monitored since the phosphoethanolamine head group is lost as a neutral fragment [1,11–13].

As these MS-based approaches become increasingly favored, data interpretation is becoming a bottleneck of lipidomic studies due to the huge amounts of data produced and the inherent complexity of biological samples, which requires extensive manual work. There are only a few computational tools available to assist the rapid and automated lipidomic data analysis process. Current computational tools can be generally classified into two groups according to the lipidomic platforms used, *i.e.*, MS profiling group and LC–MS group. The former is based on shotgun approach only, where different algorithms are specifically designed for different types of datasets acquired by various MS instruments [14–18]. Compared to the MS profiling group, few algorithms are available for LC–MS data even though the LC–MS approach has some substantial advantages over shotgun MS concerning accurate identification, quantitation, reduced ion suppression and assignment of acyl chains [19]. The available LC–MS based algorithms have various functions directing at different types of molecules, but few take full advantage of the chromatographic separation since they are either based on extracting the mass spectra as input or on extracting chromatograms for identification and quantification [20–23]. In general, the available algorithms do not provide an efficient solution for fully utilizing the useful chromatographic and mass spectral information derived from the LC–MS based data.

In the present study, a novel approach is proposed for spectral deconvolution, identification and quantification of the molecular PC and PE species in LC–MS data acquired by PIS and NLS modes. The approach is based on calculation of theoretical mass spectra of possible PC and PE compounds from a list of fatty acids followed by resolution of the sum spectrum and the full chromatographic profile by least squares regression. Since the theoretical spectra contain the isotope pattern of the individual compounds, there is no need for additional isotope correction. The principle and the algorithm are described in the theory section below. The performance of the developed algorithm was tested by using both standard mixtures and extracts of brain lipids and the results demonstrate that it is an efficient tool in lipidomic data analysis.

2. Theory

2.1. Nomenclature

In the following theory section and elsewhere in the manuscript, standard matrix notation is applied, where scalar values are denoted by lower case letters in italics, vectors are denoted by bold lower case letters and matrices are denoted by bold uppercase letters. Superscript 'T' on a matrix indicates that it is transposed, while superscript 'T' on a vector denotes that it is treated as a row vector. Scalars are denoted by lower case letters in italics.

Phospholipids are referred to by the lipid class (PE or PC) followed by total number of carbon atoms and total number of double bonds in the two fatty acids in the molecule. The two hydroxy groups in the phospholipid molecules that are linked to fatty acids are denoted by the stereospecific numbering, *sn*-1 and *sn*-2 [24]. Fatty acids are referred to by total number of carbon atoms followed by the total number of double bonds, and optionally the position of the double bond counted from the methyl end of the molecule [24]. Plasmalogens are similar in structure to ordinary PC and PE, but with a vinyl-ether linked hydrocarbon chain in the *sn*-1 position. Plasmalogens are denoted by the total number of carbon atoms followed by total number of double bonds in the fatty acid and hydrocarbon chain, excluding the double bond in the vinyl-ether group. Sphingomyelins (SM) are molecules based on a sphingoid base where a fatty acid is bound to the amino-group in the 2-position via an amide bond, and phosphocholine is linked to the hydroxy-group in the 1-position. SM are denoted by the total number of carbon atoms followed by total number of double bonds in the fatty acid and the sphingoid base.

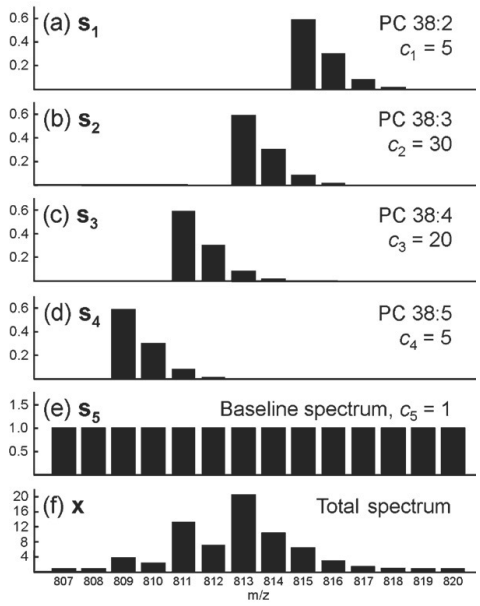


Fig. 1. Simulated unit-resolution PIS mass spectra (a) PC 38:2, (b) PC 38:3, (c) PC 38:4, (d) PC 38:5, (e) baseline spectrum and (f) total spectrum generated from (a)–(e). See Section 2.1 for lipid nomenclature.

2.2. Principle

Fig. 1a–d shows simulated unit-resolution PIS mass spectra of PC species with 38 carbon atoms and 2–5 double bonds in the fatty acid chains. The lightest ions in each spectrum, corresponding to the nominal mass of $[MH]^+$, constitute approximately 59% of the signal, while the masses of $[MH+1]^+$, $[MH+2]^+$, $[MH+3]^+$ and $[MH+4]^+$ account for approximately 30%, 8.6%, 1.7% and 0.2%, respectively. These ratios will vary slightly with molecular composition. Fatty acids typically have from 12 to 24 carbon atoms and the resulting PC molecules will therefore have from 24 to 48 carbon atoms. $[MH]^+$ will account for 69% and 53% of the signal in PC 24:0 and 48:0, respectively.

It can be seen from the figure that there is overlap between the isotope patterns of the different PC species. The $[MH]^+$ signal will be influenced by $[MH+2]^+$ from a compound with one more double bond, and the $[MH+2]^+$ signal will be influenced by the $[MH]^+$ signal from a compound with one less double bond. $[MH+3]^+$ and $[MH+4]^+$ also interfere with other signals, but these are of less importance. Different isotope correction strategies have been proposed to correct for the overlap of isotope patterns [16,20,23].

In addition to the contribution from the different molecular species, a mass spectrum may also have a baseline resulting from noise contributions, especially if it is averaged from a large number

of scans. This is illustrated in Fig. 1e, where the baseline is equal on all masses.

When a total mass spectrum, $\mathbf{x}^T$, has contributions from several compounds, the signal is a result of the concentration of the individual compounds, c, multiplied with the corresponding normalized mass spectra, $\mathbf{s}^T$, as shown in Eq. (1).

$$\mathbf{x}^T = \mathbf{s}_1^T \cdot c_1 + \mathbf{s}_2^T \cdot c_2 + \cdots + \mathbf{s}_n^T \cdot c_n \quad (1)$$

The concentrations, c, will be 'pseudo concentrations', since different compounds will have different response factors and effects such as ion suppression may lead to synergistic interactions between the different compounds. For the sake of simplicity, c is referred to as concentration, with reference to the measured signal. The hypothetical spectrum $\mathbf{x}^T$ generated from the individual spectra and concentrations varying from 5 to 30, and with a baseline level of 1, is shown in Fig. 1f.

If the relative distributions of the masses in the spectra are known, it is possible to find the concentrations of the individual compounds, c, from the total spectrum, $\mathbf{x}^T$, by least squares regression according to Eq. (2).

$$\mathbf{c} = \mathbf{x}^T\mathbf{S}(\mathbf{S}^T\mathbf{S})^{-1} \quad (2)$$

where $\mathbf{S}^T$ is an $m \times n$ matrix holding the m individual spectra of n ions, and $\mathbf{c}$ will be a vector of the m concentrations. As long as the individual spectra in $\mathbf{S}^T$ are normalized to sum = 1 (L1-normalization), the regression coefficients in $\mathbf{c}$ will be a direct estimate of the total signal from the corresponding compounds. When the baseline spectrum is described by a vector of ones, the regression coefficient for the baseline will be a direct estimate of the average baseline level of each individual ion.

In a case where all spectra in $\mathbf{S}^T$ are unique and none of the spectra can be explained by linear combination of other spectra in $\mathbf{S}^T$, Eq. (2) will give a unique solution from $\mathbf{x}^T$ and $\mathbf{S}^T$ if there is no noise in the system. These criteria are fulfilled for PIS and NLS spectra of PC and PE as long as the fatty acids are even-numbered and there are less than 14 double bonds in the two fatty acid chains. Docosahexaenoic acid (22:6 n-3), which is the most unsaturated common fatty acid, has six double bonds, so the maximum total number of double bonds from common fatty acids in PC and PE molecules will be 12. However, if fatty acids with odd-numbered carbon chains are included there will be spectral similarities between compounds differing by one carbon and seven double bonds. Due to a different number of carbon atoms and hydrogen atoms, the spectra will not be identical, but the difference between them will be less than the noise that can be expected in any real system. In principle, the methodology described above is not limited to unit-resolution mass spectra, and a mass resolution of 0.1 Da or better would solve the problem.

2.3. Algorithm

A spectral resolution based on the principle described above will eliminate the need for isotope correction and baseline subtraction. Based on this principle the following algorithm for identification and quantification of PC and PE analyzed respectively by PIS and NLS was developed:

2.3.1. Calculation of sum spectrum $\boldsymbol{x}^T$

A sum spectrum, $\mathbf{x}^T$ is created from the entire raw data set or from a sub-section of an LC–MS chromatogram.

2.3.2. Calculation of possible compounds

From a list of common fatty acids molecular compositions of possible PC or PE species are calculated. In the same step, the equivalent carbon numbers (ECN) of the species are calculated. An option at this stage is to constrain the list of compounds to a certain range of ECN, which would be natural when $\mathbf{x}^T$ is from a sub-section of a chromatogram. At this stage it is also possible to add compounds in addition to the ordinary PC or PE species, such as other lipid classes or artifacts. The molecular composition of these lipids classes can be calculated from fatty acids, or molecules can be specified individually. In this work sphingomyelins was for instance added in one of the examples shown in the results section.

2.3.3. Generation of theoretical mass spectra, $\boldsymbol{S}^T$

From the molecular composition of step 2, the theoretical mass spectra resembling the isotope distributions are calculated and normalized to sum = 1. The spectra are calculated with a mass accuracy of four decimals and thereafter binned to a selected resolution (1 in this case). The spectra are stored in $\mathbf{S}^T$. Before the binning to unit resolution, it may be necessary to add a mass offset to the calculated spectra. With no offset, $[MH]^+$ of PC species with less than 28 carbon atoms and PE species with less than 31 carbon atoms will be rounded to even masses, while the other compounds will be rounded to odd masses. An offset of +0.1 Da for PC and +0.2 Da for PE will ensure that all compounds are rounded to odd masses, and an offset of −0.4 Da for PC and −0.3 Da for PE will ensure that all are rounded to even masses. Whether positive or negative offsets are used is irrelevant for the algorithm as long as the experimental data in $\mathbf{x}^T$ is treated in a similar way.

2.3.4. Deleting absent spectra from $\boldsymbol{S}^T$

Spectra in $\mathbf{S}^T$ where the most abundant ion ($[MH]^+$ in this case) is not present in $\mathbf{x}^T$ above a predefined threshold are regarded absent from the total spectrum. These are therefore deleted from $\mathbf{S}^T$.

2.3.5. Filter for spectral similarity

The remaining spectra in $\mathbf{S}^T$ are then checked for spectral similarity. As explained above, this filter is only necessary if fatty acids with odd number of carbon atoms are included in step 2. If the correlation coefficient between two spectra are above 0.9 and the two spectra stem from compounds with even and odd number of fatty acid carbon atoms, the spectrum from the compound with odd number of carbon atoms is deleted from $\mathbf{S}^T$. In this way, spectra that can only be explained by compounds having an odd number of carbon atoms are kept.

2.3.6. Addition of baseline

A baseline spectrum, which is a row vector of ones with the same number of columns as in $\mathbf{S}^T$ is added as the last row in $\mathbf{S}^T$. This step can be omitted if $\mathbf{x}^T$ is from baseline subtracted spectra or if the baseline for other reasons is insignificant.

2.3.7. Resolution

Prior to resolution, both $\mathbf{S}$ and $\mathbf{x}^T$ are constrained to contain only masses that are present in both. The concentrations of each of the compounds remaining in $\mathbf{S}^T$ are thereafter calculated by Eq. (2). When there is noise in $\mathbf{x}^T$, a large number of minor compounds may be found, and compounds may also have small negative values in $\mathbf{c}$. For the best results, it may therefore be necessary to do a recalculation after constraining $\mathbf{S}^T$ to compounds with concentrations above a certain level.

2.3.8. Validation

A validation spectrum $\mathbf{x}_{val}^T$ can be calculated from $\mathbf{S}^T$ and the calculated $\mathbf{c}$ according to Eq. (3).

$$\mathbf{x}_{val}^T = \mathbf{c} \times \mathbf{S}^T \quad (3)$$

$\mathbf{x}^T$ and $\mathbf{x}_{val}^T$ are thereafter compared. Any deviations between the two indicate that the resolution is inaccurate. Any ions that were above the threshold in step 4 and that are not among the ions in $\mathbf{S}^T$,

also indicate that there are compounds contributing to the original $\mathbf{x}^T$ that are not accounted for.

2.3.9. Chromatographic resolution

In cases where $\mathbf{x}^T$ stems from a chromatogram or from a section of a chromatogram, the results can also be validated by chromatographic resolution according to Eq. (4), which is the central equation in most multivariate techniques for chromatographic deconvolution based on spectral information [25].

$$\mathbf{C} = \mathbf{X}\mathbf{S}(\mathbf{S}^T\mathbf{S})^{-1} \quad (4)$$

This equation is similar to Eq. (2), except that $\mathbf{X}$ is now a matrix of the two-dimensional raw data with number of rows corresponding to the number of scans (retention times) and number of columns corresponding to the number of ions in the mass spectra. $\mathbf{C}$ will then be a matrix holding the chromatographic profiles of each compound in $\mathbf{S}^T$ with number of rows corresponding to the number of scans and number of columns corresponding to the number of compounds. The resolved chromatographic profiles can be used to check that compounds with similar ECN values elute in the same regions (in reversed-phase LC).

Similar to the resolution by Eq. (2), a flat baseline spectrum can be added to $\mathbf{S}^T$ in Eq. (4). Any structure in the corresponding chromatographic profile indicates that there are peaks that are not accounted for. It can also be checked that the theoretical spectra, $\mathbf{S}^T$, and the resolved chromatograms, $\mathbf{C}$, explain the structure in $\mathbf{X}$ by calculating residuals according to Eq. (5).

$$\mathbf{R} = \mathbf{X} - \mathbf{C}\mathbf{S}^T \quad (5)$$

Several fatty acid combinations may give PC and PE species with the same number of carbon atoms and double bonds, and the chromatograms in $\mathbf{C}$ may also be used to quantify these if there is sufficient chromatographic resolution.

3. Experimental

3.1. Standards and samples

PC and PE standards, including 1,2-didocosahexaenoyl-*sn*-glycero-3-phosphocholine (PC 22:6/22:6), 1-palmitoyl-2-docosahexaenoyl-*sn*-glycero-3-phosphocholine (PC 16:0/22:6), 1-palmitoyl-2-arachidonoyl-*sn*-glycero-3-phosphocholine (PC 16:0/20:4), 1-stearoyl-2-docosa-hexaenoyl-*sn*-glycero-3-phosphocholine (PC 18:0/22:6), 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine (PC 16:0/18:1), 1-octadecanoyl-2-(5Z,8Z,11Z,14Z,17Z-eicosapentaenoyl)-*sn*-glycero-3-phosphocholine (PC 18:0/20:5), 1-palmitoyl-2-docosahexaenoyl-*sn*-glycero-3-phosphoethanolamine (PE 16:0/22:6), 1,2-didocosahexaenoyl-*sn*-glycero-3-phosphoethanol-amine (PE 22:6/22:6) and 1-stearoyl-2-docosahexaenoyl-*sn*-glycero-3-phosphoethanolamine (PE 18:0/22:6), and porcine brain sphingomyelins were purchased from Avanti Polar Lipids, Inc. The stock solutions of these standards were prepared in chloroform. Two standard mixtures containing 10 μg/mL of the six PC species and the three PE species were prepared by dilution in methanol/acetonitrile (60:40, v/v). HPLC grade acetonitrile and LC–MS grade methanol were purchased from Merck (Darmstadt, Germany) and Sigma–Aldrich (St Louis, MO, USA), respectively.

Total lipids were extracted from mouse and cod brain samples by the Folch principle [26]. Briefly, samples were homogenized in chloroform/methanol (2:1, v/v), thoroughly vortex-mixed and filtered on Büchner funnel. Cod Brain extracts were washed with 0.2 volumes of 0.88% KCl and the water phase was discarded. The final extracts were evaporated to dryness and dissolved in 2 mL chloroform. An aliquot of 500 μL brain sample was taken out from this extract and dissolved in 500 μL methanol/acetonitrile (60:40, v/v) for the analysis of LC–MS.

3.2. LC–MS analysis

LC–MS analyses were carried out by using a 1100 series LC (Agilent) with a binary pump, autosampler, thermostatted column compartment connected to a Micromass Quattro MS equipped with an electrospray interface (Waters, Manchester, UK). The system was controlled by the MassLynx Software (Waters).

A reversed phase 50 mm × 2.1 mm ACE 5 C_{18} column with 5 μm particles (Advanced Chromatography Technologies, Aberdeen, UK), was used for the separations, and the column was kept at 35 °C. The LC conditions were adapted from [27]. Methanol/acetonitrile/water, 45:30:25 by volume, (Solvent A) and methanol/acetonitrile, 60:40 by volume (Solvent B) were used as mobile phases and both Solvent A and B contained 2.5 mM ammonium acetate (Sigma–Aldrich) and 10 μM serine (Sigma–Aldrich). The gradient program started at 40% B and increased to 100% B in 15 min, held for 45 min and returned to the initial condition in 1 min. The equilibration time between the injections was 9 min.

The ion source parameters were optimized by using both PC and PE standards. The ion source operated in positive mode, the source temperature was set to 150 °C and desolvation temperature was 400 °C. Nitrogen was used as curtain gas and argon was used as collision gas. PC species were detected by using PIS for m/z 184 with the collision energy of 45 eV and scanning for m/z 620 to 965, while PE species were monitored by using NLS for m/z 141 with the collision energy of 35 eV, and scanning in the range of m/z 575 to 920. The LC–MS raw data were exported to the self-describing, machine-independent data format, NetCDF (www.unidata.ucar.edu) by the DataBridge tool of MassLynx 4.0 Software.

3.3. Fatty acid analyses

Fatty acid methyl esters were prepared by as described in [28] and analyzed by GC-FID. Mouse brain samples were analyzed on a BPX-70 (SGE, Ringwood, Australia) using conditions described in [29] and cod brain samples were analyzed on a BP-20 column (SGE) using conditions described in [28]. In both cases empirical response factors calculated from the reference mixtures GLC463 or GLC-793 (Nu-Chek prep, Elysian, MN, USA) were applied for correcting the chromatographic areas, and results were expressed as percent of total fatty acids.

3.4. Software

The described algorithm was implemented in an in-house written program, Chrombox D 12-09 (www.chrombox.org) running under Matlab (Natick, MA, USA). NetCDF files with LC–MS raw data were read into Chrombox D and masses were binned to unit resolution with mass offsets of +0.2 Da and −0.1 Da for PC and PE data respectively. This ensured that the masses in the PC spectra were rounded upwards and masses of PE spectra were rounded downwards. Theoretical spectra were calculated with high resolution (4 decimals) in step 3 of the algorithm and then binned to unit resolution with offsets of +0.2 Da and −0.3 Da for PC and PE, respectively.

The following list of fatty acids was applied in step 2 of the algorithm: 12:0, 14:0, 14:1, 15:0, 16:0, 16:1, 17:0, 17:1, 18:0, 18:1, 18:2, 18:3, 18:4, 19:0, 19:1, 20:0, 20:1, 20:2, 20:3, 20:4, 20:5, 22:0, 22:1, 22:2, 22:3, 22:4, 22:5, 22:6, 24:0 and 24:1. This list of fatty acids will calculate 133 possible compounds in each lipid class, varying in number of carbon atoms from 24 to 48 and total number of double bonds from 0 to 12 in the two acyl chains. The threshold described in step 4 of the algorithm was set to 2% of the base peak. Since the

list of fatty acids contains fatty acids with odd-numbered carbon chains, the filter in step 5 was active in the algorithm. A flat spectral baseline was added to the spectra in $\mathbf{S}^T$, as described for steps 6 and 9 in the algorithm. After the first calculation the results were recalculated excluding compounds with less than 3% relative to the most abundant PC molecule, or less than 2% of the most abundant PE molecule. In the case of PC analysis in mouse brain, theoretical spectra of PC based SM were also calculated together with ordinary PC in step 2 of the algorithm. The ranges of chain lengths and number of double bonds in both the sphingoid bases and the fatty acids in the SM were the same as specified above.

4. Results and discussion

4.1. Analysis of PC in cod brain

The performance of the algorithm was tested by reference mixtures as well as biological samples such as cod brain and mouse brain. Results for the reference mixtures are given in supplementary material. Results for the brain extracts are described and discussed below. The total spectrum shown in Fig. 2a was extracted from the retention time range 15–46 min. The estimated baseline of the total spectrum is marked by a horizontal green line. The threshold value of the baseline described in step 4 in the algorithm is marked by the horizontal red line. Green bars are the masses in common with theoretical masses that remain in $\mathbf{S}^T$ after step 4, while red bars are masses that were above the threshold but not present in the theoretical spectra calculated in step 3. Other masses are shown in blue.

As can be seen in Fig. 2a, the total mass spectrum is quite complicated with the majority of the signals in the mass region m/z 700–900. Only a few ions above the threshold value were not accounted for by the spectra in $\mathbf{S}^T$. All of these were of minor abundance. With the complexity of the total mass spectrum there is significant overlap of the isotopic patterns of the individual compounds that is handled by the regression as explained in the theory section.

The amounts of the different compounds after resolution are shown by the bar plot in Fig. 2b. This is a plot of the values in $\mathbf{c}$ after application of Eq. (2), excluding the baseline signal that is also present in $\mathbf{c}$. The numbers in brackets are the ECN values of the compounds. The bar plot reveals that the main abundant PC species are PC 34:1, PC 36:1, PC 38:6 and PC 42:2, which are probably constituted by the following fatty acyl combinations, *i.e.*, 16:0/18:1, 18:0/18:1 or 16:0/20:1, 16:0/22:6 or 18:1/20:5 and 18:1/24:1, respectively. A full list of the identified compounds with possible fatty acid combinations are given as supplementary material in Table S1. These results are in close accordance with those reported previously [30–32], where the fatty acids or PC species compositions from cod brain were studied. The fatty acid combination also fits well with the fatty acid composition of the extracts obtained from the GC analysis (Table 1), which shows that 18:1, 22:6, 16:0, 18:0, 24:1, 20:5, 20:1 are the major fatty acids in the extracts. However, it is emphasized that the results in Table 1 are the fatty acid composition in the crude extract and that the composition in the different lipid classes will vary [30,32].

The result from the validation (algorithm step 8) is illustrated in Fig. 2c, where the abundances in $\mathbf{x}^T_{val}$ (Eq. (3)) are plotted against the corresponding masses from the original $\mathbf{x}^T$ shown in Fig. 2a. It can be seen that the original sum spectrum is accounted for with high accuracy by the spectrum calculated from the regression coefficients in $\mathbf{c}$ and the theoretical spectra in $\mathbf{S}^T$.

The result from the chromatographic resolution, as explained in step 9 of the algorithm gives additional information about the identity of the compounds. The data matrix $\mathbf{X}$, resolved profiles

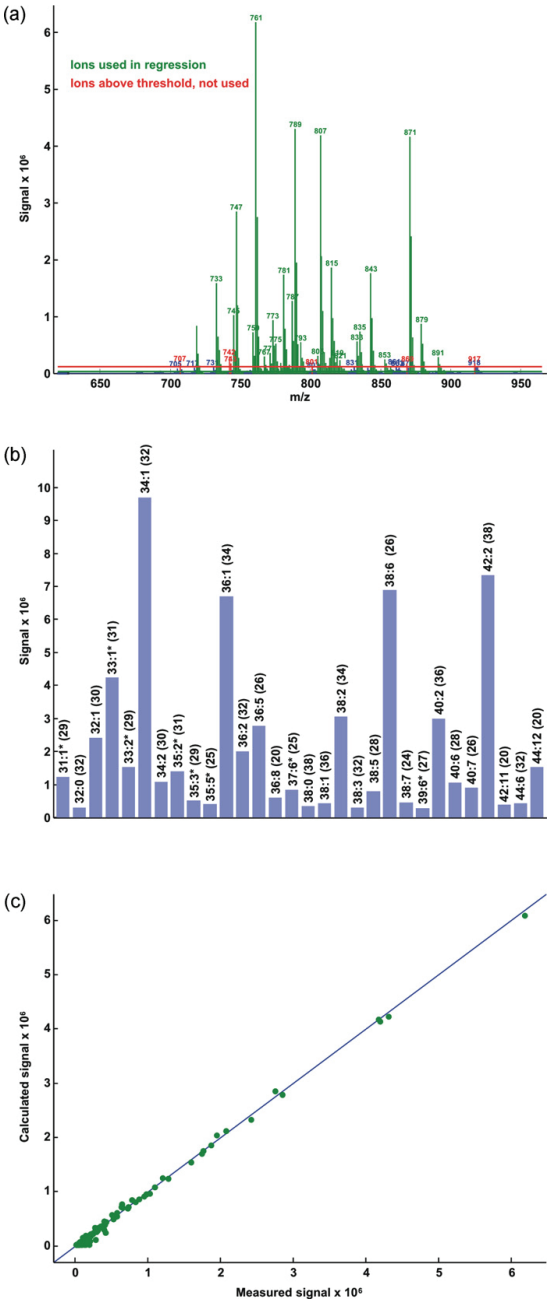


Fig. 2. Analysis of PC species from cod brain extracts. (a) The sum spectrum. Green bars are the masses used by the algorithm, red bars are other masses above a threshold marked by the horizontal red line, the horizontal green line is the estimated baseline; (b) Identified PC species with their pseudo concentrations showed by *y*-axis. Numbers in brackets are ECN values; (c) Validation of the results (predicted versus measured when the green masses in the total spectrum was reconstructed from the calculated solution). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

Table 1
Mass fraction of fatty acids (g/100 g) in cod brain and mouse brain samples. Fatty acids that were above 0.1% in one of the samples are reported.

	Cod brain	Mouse brain
14:0	0.53	0.15
16:0	14.42	23.28
16:1[a]	3.17	0.94
17:0	0.35	0.12
17:1	0.58	<0.01
18:0	9.22	20.83
18:1[b]	27.01	17.76
18:2 n-6	0.33	0.55
20:0	0.14	0.25
20:1[b]	3.32	1.17
20:2 n-6	0.19	0.16
20:3 n-6	0.07	0.42
20:4 n-3	0.27	<0.01
20:4 n-6	0.66	10.71
20:5 n-3	6.44	<0.01
22:0	0.16	0.31
22:1[b]	1.52	<0.01
22:2 n-6	0.10	<0.01
22:4 n-6	0.15	2.56
22:5 n-3	0.95	<0.01
22:6 n-3	21.26	19.12
24:0	0.37	0.51
24:1[b]	8.58	1.16

[a] Basically n-7 isomer.
[b] Basically n-9 isomer.

C (Eq. (4)) and residuals **R** (Eq. (5)) are shown on equal scales in Fig. 3a–c, respectively. The signals in **C** are larger than the signals in **X** because **C** contains the sum of all isotopomers of each PC species. Irregularly shaped peaks or double peaks should be expected since several combinations of fatty acids may give PC species with the same masses.

The residuals are small compared to the magnitudes in **C** and **X**, showing that the majority of variance in **X** has been accounted for in all parts of the chromatograms. The majority of the variance in **X** are spikes with short frequencies in the regions where there are chromatographic peaks, probably arising from heteroscedastic noise in **X**. But some of the ions in **R** show similarity to chromatographic profiles, such as m/z 779, 783 and 814, indicating that there may be minor compounds that are not accounted for by the spectra in $\mathbf{S}^T$.

The retention of PC species in reverse phase LC follows their ECN values due to specific hydrophobic interaction of their fatty acyl chains with the alkyl ligand of the stationary phase [33]. This means that the PC species having the same ECN values are basically in the same retention time region and the PC species are gradually eluted with increasing ECN values. This rule can be used to assist the identification of the compounds since false identifications based on the spectra alone may arise from the presence of minor compounds such as ether-linked plasmanyl or plasmenyl PC species (plasmalogens). Most of the ether-linked species can be distinguished from the diacyl PC species through the bar plot (Fig. 2b) or the resolved chromatographic profiles (Fig. 3b). Common plasmalogens has the same integer masses as ordinary PC with odd number of carbon atoms, and it has been shown that ether-linked PC species elute later than their diacyl counterparts on a reverse phase column [7,34].

The most abundant compounds identified as PC with odd number of carbon atoms are 33:1, 35:2, 31:1 and 33:2, which have the same integer masses as 34:0, 36:1, 32:0 and 34:1 plasmalogens, respectively. These are among the most common and abundant PC plasmalogens previously reported in brain [35]. They also elute later than what should be expected from their calculated ECN number, which should be expected according to [7,34] if they are plasmalogens.

There are two peaks corresponding to PC 33:2, where one has the expected retention time for an ordinary PC with odd number of carbon atoms, and one peak has stronger retention indicating that it is the 34:1 plasmalogen. PC 39:6 also has correct ECN value for ordinary PC and can be explained by the fatty acid combination 17:0 and 22:6 (Table 1). The remaining compounds reported as PC with odd number of carbon atoms (37:6, 35:3, 35:5) can either be explained by previously reported plasmalogens in brain [35] or by the abundant 22:6 or 22:5 fatty acids combined with ether linked C16 or C18 alkenyl or alkanyl chains.

4.2. Analysis of PE in cod brain

The results for the PE analysis are presented in Figs. 4 and 5. The extracted total mass spectrum from the retention time range between 16 and 31 min (Fig. 4a) are simpler than the corresponding spectrum for PC.

Most of the ions are concentrated in the mass region m/z 700–850 and a few minor ions above the threshold value marked by red were not accounted in the spectrum $\mathbf{S}^T$. The most abundant ions, m/z 836, 764, 792 and 790 correspond to the base peaks of PE 44:12, PE 38:6, PE 40:6, PE 40:7, respectively, Fig. 4b. Based on the fatty acid composition in Table 1, it is likely that the dominating PE species have the following fatty acyl combinations: 22:6/22:6 for PE 44:12, 16:0/22:6 or 18:1/20:5 for PE 38:6, 18:0/22:6, 18:1/22:5 or 20:1/20:5 for PE 40:6 and 18:1/22:6 for PE 40:7. A full list of proposed structures is given in the supplementary material, Table S2.

This presence of mainly 22:6, 18:1, 16:0 and 18:0 containing PE species coincides well with other studies [30–32]. The result from the validation (algorithm step 8) shown in Fig. 4c reveals that the deviation between the predicted and measured total spectrum is small, and there are only a few minor masses contributing to the total signal that are not accounted for.

The results from the chromatographic resolution after the application of algorithm step 9 are given in Fig. 5. Similar to PC, the elution pattern of diacyl PE and ether-linked PE species can be used as a diagnostic tool for the accurate identification of these species. It is worth noting that the resolved chromatogram (Fig. 5b) indicates the presence of several isobaric species which have the same mass but different fatty acyl combinations. PE 38:6 has two closely eluting chromatographic peaks, possibly consisting of 16:0/22:6 and 18:1/20:5. Similarly, PE 40:6 consists of three chromatographic peaks which probably correspond to 18:0/22:6, 18:1/22:5 and 20:1/20:5. The resolved PE chromatogram (Fig. 5b) only show four minor peaks with odd number of carbon atoms that may arise from ether-linked species [36].

4.3. Mammalian brain lipids

The bar plots from the mouse brain analysis are presented in Fig. 6 and the proposed identity of the compounds are listed in the supplementary material, Tables S3 and S4. The main mouse brain PC species were 34:1, 36:1, 32:0 and 38:6, constituted by 16:0/18:1, 18:0/18:1 or 16:0/20:1, 16:0/16:0 and 16:0/22:6, respectively (Fig. 6a). The same four compounds have previously been reported to be the most abundant PC species in mouse brain [37]. Although the mouse brain PC profile is less complex than the corresponding cod brain profile, they share several abundant species, such as PC 34:1 (16:0/18:1), 36:1 (18:0/18:1 or 16:0/20:1) and 38:6 (16:0/22:6).

The results for mouse brain PC were calculated with the function for SM active in step 2 of the algorithm, in addition to the function for ordinary PC. Because SM have different number of N atoms than ordinary PC they cannot have similar spectra. They are therefore less problematic to analyze together with PC than for instance

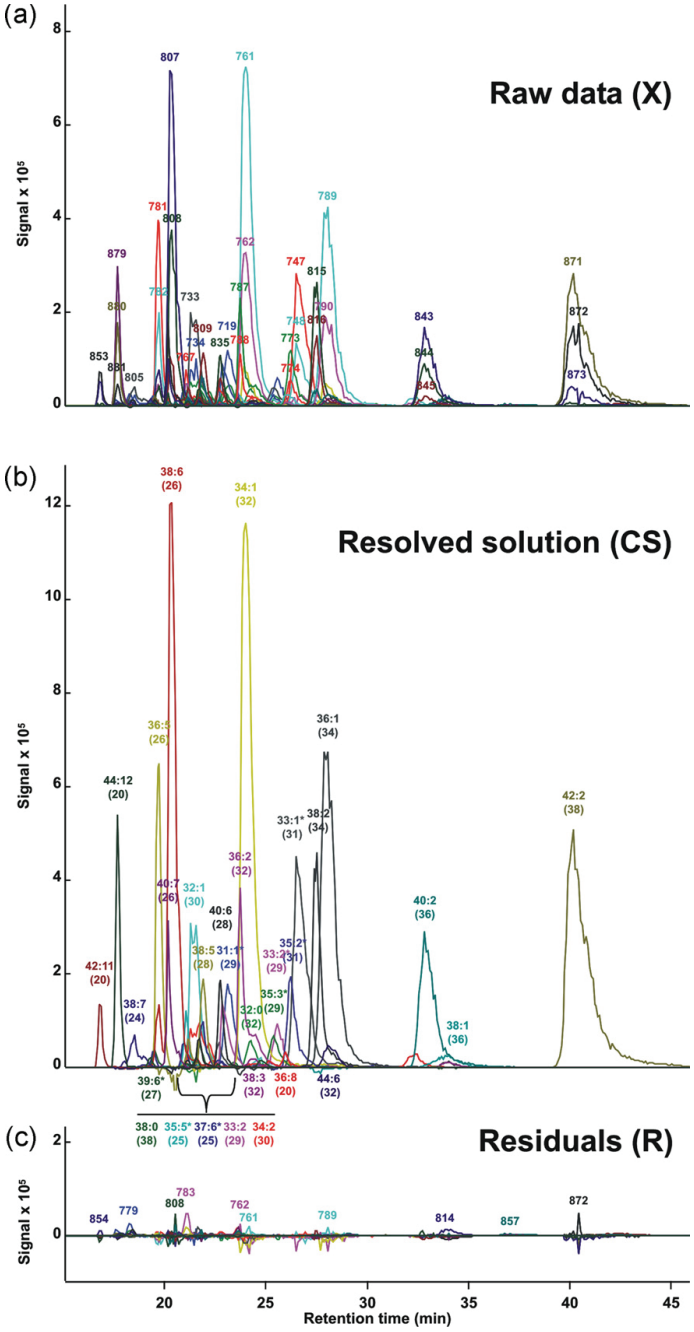


Fig. 3. The plot of raw data matrix **X** (a), resolved chromatographic profiles **C** (b) and residuals **R** after resolution (c) of PC species from cod brain extracts. (a–c) are shown on equal scales.

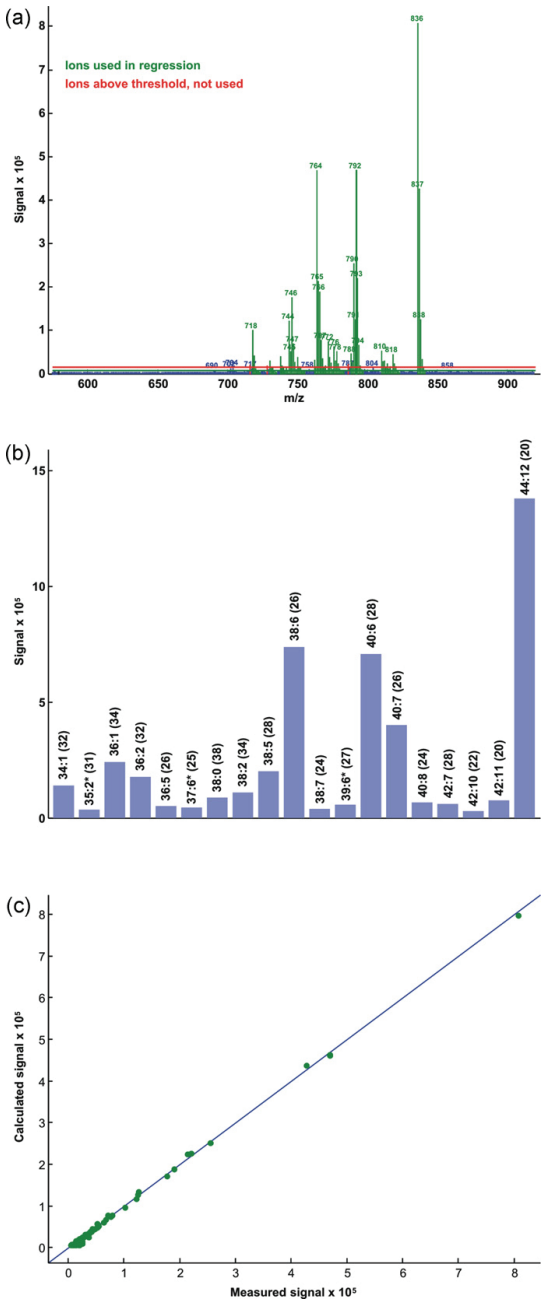


Fig. 4. Analysis of PE species from cod brain extracts. (a) The sum spectrum. Green bars are the masses used by the algorithm, red bars are other masses above a threshold marked by the horizontal red line, the horizontal green line is the estimated baseline; (b) Identified PE species with their pseudo concentrations showed by *y*-axis. Numbers in brackets are ECN values; (c) Validation of the results (predicted versus measured when the green masses in the total spectrum was reconstructed from the calculated solution). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

the plasmalogens. Only three compounds were detected above the threshold of the 133 theoretically possible SM molecular formulas that were calculated from the list of fatty acids. These were SM 36:1, SM 36:2 and SM 42:2. All three have previously been reported to be among the most abundant SM in mammalian brain with 36:1 and 42:2 as the two most abundant [38,39].

It should be emphasized that our extraction methods and LC–MS parameters are not optimized for SM, so the levels shown in Fig. 6a may not be representative for the ratio between ordinary SM and ordinary PC. The sum of SM molecules is approximately 7% of the sum of PC. Ratios of SM to PC in mouse brain have been reported to be typically around 10% [40].

Results for an analysis of a porcine brain SM reference mixture analyzed by the same method are reported as supplementary material (Table S5 and Figures S5 and S6). No ordinary PC was detected in this sample. The levels of different SM agreed well with a previous report on the bovine brain SM [39]. Two of the detected compounds have deviating ECN values indicating that they are not ordinary SM or PC. Of the remaining 14 compounds, 12 were reported in the previous work [39], and only one of the abundant compounds (>1% relative to largest) reported in this work was not detected above the threshold (SM 43:1).

The bar plot of mouse brain PE (Fig. 6b) suggests that the main species are 40:6, 38:4, 38:6 and 36:1, constituted by 18:0/22:6, 18:0/20:4 or 16:0/22:4, 16:0/22:6 and 18:0/18:1 or 16:0/20:1, respectively. The main trends in the obtained results are in accordance with previous works [41–43].

In the fatty acids profiles of the extracts (Table 1), the most striking difference between cod and mouse brains can be seen in the levels of the polyunsaturated fatty acids 20:4 and 20:5, and partly also in the elongated 22:4 and 22:5. While 20:5 and 22:5 are high in cod brain these are absent in the mouse brain, and 20:4 and 22:4 that are found in high levels in mouse brain are in very low levels in the cod brain. This is also reflected in the results found from the LC–MS analyses. Except from PE 42:10 (20:4/22:6), which was found in minor amounts in cod brain, 20:4 and 22:4 seem absent in PC and PE from cod brain. Similarly, 20:5 and 22:5 are absent in the mouse brain PE and PC.

4.4. Discussion of the algorithm

The algorithm as presented in Section 2.2 can be extended or modified in several ways. One obvious way to adapt the procedure to different types of samples is to modify the list of fatty acids to contain only fatty acids that are present in the samples above a certain threshold. Step 2 in the algorithm can generate spectra of more than one lipid class, which was used in the analysis of mouse brain PC. In addition to specifying sphingolipids in cases where this is relevant, one can also include spectra of other compounds that may be detected by the same methods as used for the main phospholipid classes, such as lyso-phospholipids or plasmalogens. However, such an extension of step 2 may also require modifications of the filter for spectral similarity in step 5. Alternatively, step 2 can be omitted and replaced by a predefined list of molecular formulas suitable to the sample type of interest. This will probably give higher numerical stability and better precision in analyses of large sequences. Similarly, both step 2 and 3 can be replaced by a predefined list of spectra.

We have only applied the methodology with unit resolution mass spectra, but in theory the procedure can be adapted also to high resolution instruments. However, instruments capable of running PIS and NLS with high resolution are few since high resolution would be required in the first mass filter.

Compared to other ionization techniques electrospray ionization usually leads to quite noisy data, which is something that can also be seen in the raw data profiles in Figs. 3a and 5a. However,

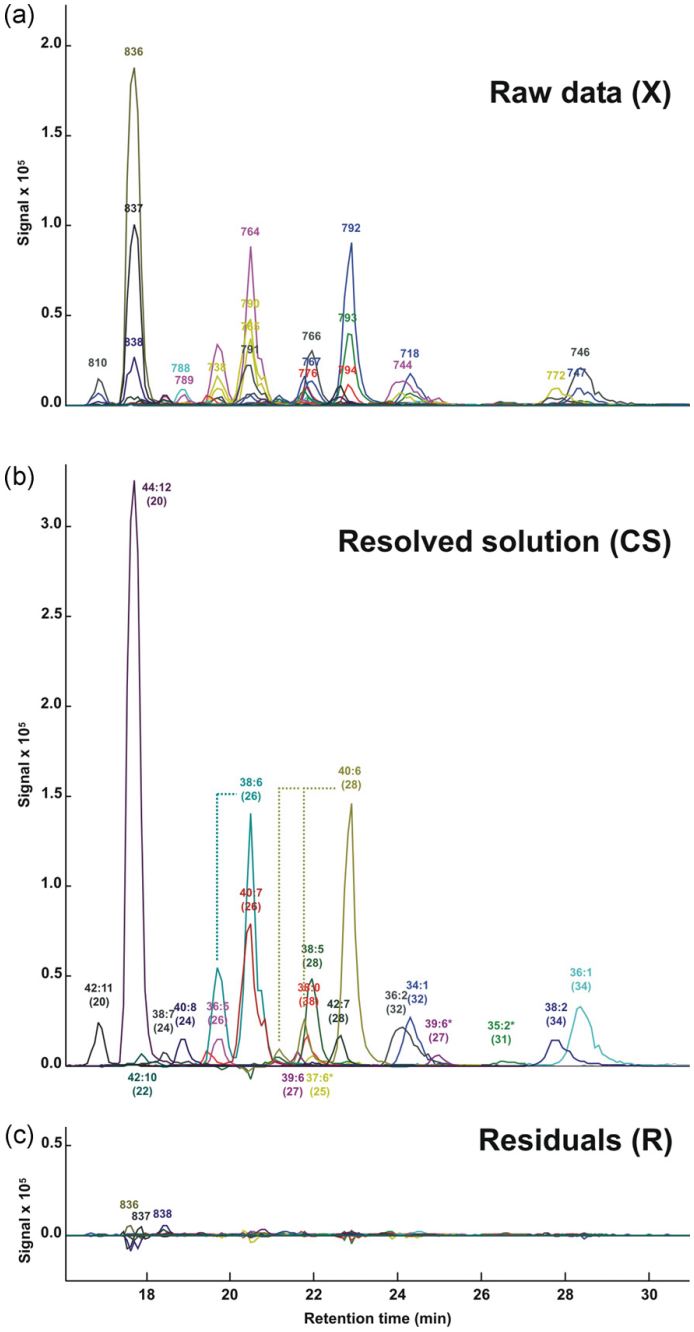


Fig. 5. The plot of raw data matrix **X** (a), resolved chromatographic profiles **C** (b) and residuals **R** after resolution (c) of PE species from cod brain extracts. (a–c) are shown on equal scales.

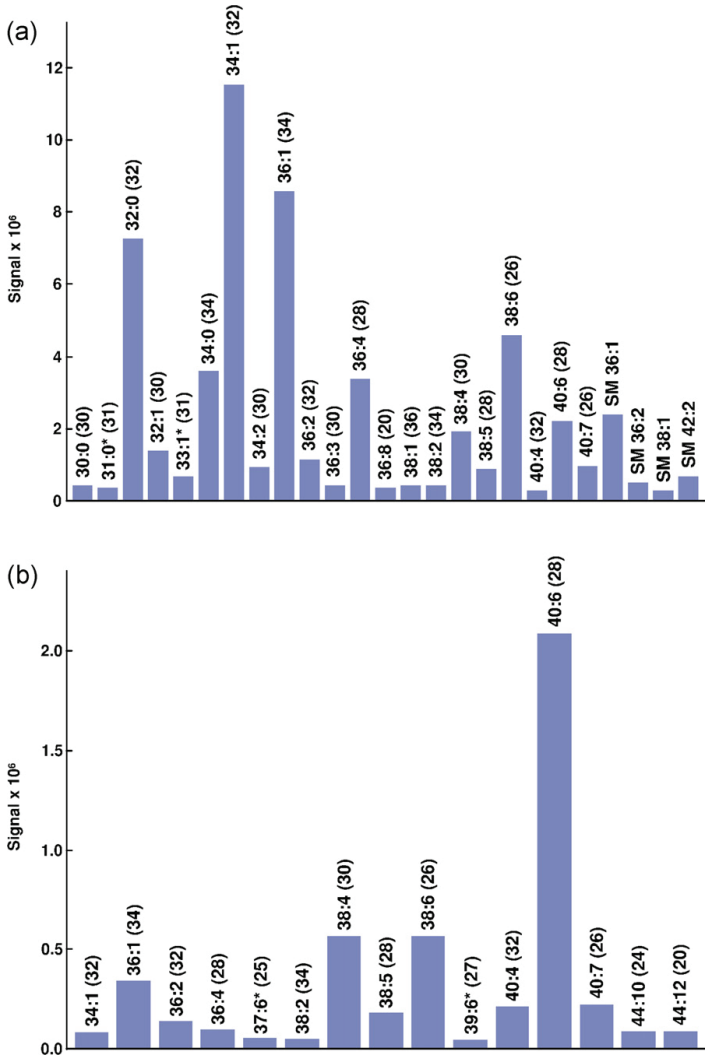


Fig. 6. Identified (a) PC, SM and (b) PE species from mouse brain extracts.

since the noise is basically caused by processes in the ion source, such as instable ionization and ion suppression, the majority of the noise in the data is added before the mass filter. With reference to Eqs. (4) and (5) one can say that the noise in **X** stems basically from **C** and not from $\mathbf{S}^{T}$. This is the reason for the quite small residuals even though the data are noisy.

The noise in mass spectrometry is usually heteroscedastic, where absolute noise in the signals increase with the signal strength [44]. This may typically affect the accuracy of quantification of minor compounds that have ions in common with major compounds in the total spectrum. This is a problem related to the structure in the raw data, so it will be equally problematic also with other methods for quantification and isotope correction that is based on a single mass spectrum. But if minor peaks are chromatographically separated from larger peaks that share some of the ions (the source of the noise) they will not be influenced in the same way. In many cases it may therefore be better to base the quantification on integrated profiles of **C** from Eq. (4) than on the regression coefficients, **c**, from Eq. (2). Alternatively, the algorithm may be used on sections of the chromatogram. In the current version of the applied program there is no possibility to integrate the profiles, but they can be exported and quantified for instance by Chrombox C (www.chrombox.org) also running under Matlab.

Noise will typically also lead to detection of a large number of minor compounds and also some compounds with small negative values in **c**, depending on the threshold selected in step 4. These can be removed by recalculating only with compounds that had signal above a certain value. However, since the signals from compounds that are omitted from the recalculation are still present in the

measured spectrum, $\mathbf{x}^T$, they may influence accuracy of compounds that share some of the masses. Setting the threshold for the recalculation too high may therefore have negative effect on the accuracy, which will be seen in the $\mathbf{x}^T$ versus $\mathbf{x}^T_{val}$ plots. To facilitate presentation of the data, we have in this work used higher thresholds than what we would have selected if we wanted maximal accuracy for minor compounds. Another way to avoid negative values is to use a regression that constrains $\mathbf{c}$ to have only non-negative values. This is an option in the current version of the program, but it should be used with care since negative values may be of diagnostic importance.

The filtering of similar ions in step 5 of the algorithm is also implemented in the software as an option. If the filter is turned off, similar spectra will be flagged and it is possible to recalculate after manual selection of one of the similar compounds. The threshold value for the correlation between spectra of 0.9 works fine for the data presented used in this work, but it is emphasized that a suitable level may depend on the amount of noise in the data.

In theory it is possible to apply similar methodology on other compounds. But the methodology requires a reasonable fit between measured and theoretically calculated spectra, as well as limited abundance of interferents that are not accounted for by the theoretical spectra. Similar methodology has for instance been applied for deconvolution of petroleum compounds [45] and selenides [46] analyzed by high-resolution mass spectrometry.

4.5. Comparison with alternative methods

Several tools for analysis of data from direct infusion MS or LC–MS of lipids have been described in the scientific literature. These include software such as LipidProfiler [14], LipidInspector [15], LipidQA [16], Fatty acid Analysis Tool (FAAT) [17], LipidXplorer [18], LIMSA [20], MZmine2 [21], Profiler-Merger-Viewer [22] and Lipid Data Analyzer (LDA) [23]. The various programs differ widely in scope and flexibility. Some tools are basically designed for direct infusion MS [14–18], and other are designed for LC–MS [20–23]. However, programs designed for LC–MS may not take full advantage of the LC separation, and some of the direct infusion methods may handle spectra extracted from LC–MS chromatograms, so there is no clear distinction between these two classes. The methods differ in requirements for mass spectral resolution and whether centroided or full profile spectra are applied. Some methods are claimed to handle both high and low resolution MS [14–16,18,21]. The various tools also have different scopes. While some tools aim at identification and quantitation of the signal generated from each compound (like the methodology described in this paper) other have built-in modules for visualization, statistics, and absolute quantitation using standards.

The data from the experiments described in Section 3 could be analyzed by LIMSA [20] and MZmine2 [21]. Quantitative results were evaluated by comparing the 19 compounds that were found above 1% in the cod brain sample (Supplementary material, Table S1). There was a large degree of correlation with R^2 of 0.9993 and 0.9964 when the values from the LSSR algorithm were compared against LIMSA and MZmine2, respectively. However, both LIMSA and MZmine2 reported a large number of compounds (>50) that were found in very small levels. These were not present above the noise level in the raw data. The same problem was also seen with the PC reference mixture that contains only 6 compounds. This is probably caused by differences in how the background signals in the spectra are handled by the algorithms. LIMSA and MZmine also failed on identification of some of the abundant compounds. In LIMSA this problem can be solved by editing the lists that specifies the compounds that may be available in a sample. Some of the typical marine compounds containing 22:5 n-3 or 22:6 n-3 were not specified in the original peak lists. In MZmine2 there was a tendency that the compounds with highly unsaturated fatty acids were identified as compounds with odd number of carbon atoms, *e.g.* PC 44:12 was assigned as PC 43:5 that has the same mass in low resolution spectra. As explained in the theory section, this conflict is not a problem when the resolution is better than 0.1 Da. The possibility to apply a filter similar to that described as step 5 in the algorithm (Section 2.3) or any other means of giving priority to one of the conflicting compounds could make MZmine2 more suitable for low resolution spectra.

There is a large variety in software tools for lipid analysis and in the quality of the data produced on different instrumentation, and it is challenging to evaluate from published literature which tool that would give the best performance in a particular case. The only advice we would like to give on the choice between the different tools is to test them thoroughly with data from the application they are intended to be used for. Many tools specify lists of compounds that are expected to be present. Particular attention should be paid to these lists and if they cover the content of the samples. It is often possible to exclude compounds from these lists, for instance in cases where pre-separation steps or selective MS scan modes are applied, or if knowledge about the sample type tells that certain compounds cannot be present. Constraining the lists to compounds that can be expected in the samples reduces the risk of incorrect identifications and may lead to more accurate and precise quantitative results.

5. Conclusions

A methodology for automated identification and quantification of molecular forms of PC and PE analyzed by respectively neutral loss scan and precursor ion scan in electrospray ionization LC–MS has been proposed. The methodology is based on calculation of theoretical spectra of the isotope distribution of possible compounds defined by a list of fatty acids, and subsequent resolution of the raw data by least squares regression.

The described algorithm was able to resolve PC and PE species of reference mixtures, porcine brain SM and extracts of brain lipids. The built-in validation approach showed that the raw data signals were accounted for, and the obtained results were reasonable considering the fatty acid profile of the extracts and results previously reported in the literature.

The methodology can be applied on a single sum-spectrum, and may therefore be feasible for shotgun lipidomics, but resolution of chromatographic raw data may give additional information about the identity and amounts of the individual compounds. The flexibility of the algorithm allows it to be expanded to other compound classes in the future.

Acknowledgements

Mouse brain was obtained from Haukeland University Hospital and National Institute of Nutrition and Seafood Research (Bergen, Norway). Cod brain was provided by Institute of Marine Research, Bergen, Norway.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.chroma.2012.12.070.

References

[1] M. Pulfer, R.C. Murphy, Mass Spectrom. Rev. 22 (2003) 332.

[2] V. Frisardi, F. Panza, D. Seripa, T. Farooqui, A.A. Farooqui, Prog. Lipid Res. 50 (2011) 313.
[3] A.A. Farooqui, L.A. Horrocks, T. Farooqui, Chem. Phys. Lipids 106 (2000) 1.
[4] D.F. Horrobin, C.N. Bennett, Prostag. Leukotr. Ess. 60 (1999) 217.
[5] X.L. Han, R.W. Gross, Mass Spectrom. Rev. 24 (2005) 367.
[6] B.L. Peterson, B.S. Cummings, Biomed. Chromatogr. 20 (2006) 227.
[7] J.F.H.M. Brouwers, E.A.A.M. Vernooij, A.G.M. Tielens, L.M.G. van Golde, J. Lipid Res. 40 (1999) 164.
[8] H.Y. Kim, T.C.L. Wang, Y.C. Ma, Anal. Chem. 66 (1994) 3977.
[9] J.F.H.M. Brouwers, B.M. Gadella, L.M.G. van Golde, A.G.M. Tielens, J. Lipid Res. 39 (1998) 344.
[10] P.S. Niemela, S. Castillo, M. Sysi-Aho, M. Oresic, J. Chromatogr. B 877 (2009) 2855.
[11] W.D. Lehmann, M. Koester, G. Erben, D. Keppler, Anal. Biochem. 246 (1997) 102.
[12] N. Navas-Iglesias, A. Carrasco-Pancorbo, L. Cuadros-Rodriguez, TrAC-Trends Anal. Chem. 28 (2009) 393.
[13] B. Brügger, G. Erben, R. Sandhoff, F.T. Wieland, W.D. Lehmann, Proc. Natl. Acad. Sci. U.S.A. 94 (1997) 2339.
[14] C.S. Ejsing, E. Duchoslav, J. Sampaio, K. Simons, R. Bonner, C. Thiele, K. Ekroos, A. Shevchenko, Anal. Chem. 78 (2006) 6202.
[15] D. Schwudke, J. Oegema, L. Burton, E. Entchev, J.T. Hannich, C.S. Ejsing, T. Kurzchalia, A. Shevchenko, Anal. Chem. 78 (2006) 585.
[16] H. Song, F.F. Hsu, J. Ladenson, J. Turk, J. Am. Soc. Mass Spectrom. 18 (2007) 1848.
[17] M.D. Leavell, J.A. Leary, Anal. Chem. 78 (2006) 5497.
[18] R. Herzog, K. Schuhmann, D. Schwudke, J.L. Sampaio, S.R. Bornstein, M. Schroeder, A. Shevchenko, PLoS One 7 (2012) e29851.
[19] J.F. Brouwers, Biochim. Biophys. Acta 1811 (2011) 763.
[20] P. Haimi, A. Uphoff, M. Hermansson, P. Somerharju, Anal. Chem. 78 (2006) 8324.
[21] T. Pluskal, S. Castillo, A. Villar-Briones, M. Oresic, BMC Bioinformatics 11 (2010) 395.
[22] E.M. Hein, B. Bodeker, J. Nolte, H. Hayen, Rapid Commun. Mass Spectrom. 24 (2010) 2083.
[23] J. Hartler, M. Trotzmuller, C. Chitraju, F. Spener, H.C. Kofeler, G.G. Thallinger, Bioinformatics 27 (2011) 572.
[24] Anon., J. Lipid Res. 19 (1978) 114.
[25] E.R. Malinowski, Factor Analysis in Chemistry, 3rd ed., Wiley, New York, 2002.
[26] J. Folch, M. Lees, G.H.S. Stanley, J. Biol. Chem. 226 (1957) 497.
[27] K. Retra, O.B. Bleijerveld, R.A. van Gesteil, A.G.M. Tielens, J.J. van Hellemond, J.F. Brouwers, Rapid Commun. Mass Spectrom. 22 (2008) 1853.
[28] S. Meier, S.A. Mjøs, H. Joensen, O. Grahl-Nielsen, J. Chromatogr. A 1104 (2006) 291.
[29] C. Sciotto, S.A. Mjøs, Lipids 47 (2012) 659.
[30] M.V. Bell, J.R. Dick, Lipids 26 (1991) 565.
[31] S. Meier, T.C. Andersen, K. Lind-Larsen, A. Svardal, H. Holmsen, Comp. Biochem. Physiol. C 145 (2007) 420.
[32] D.R. Tocher, D.G. Harvie, Fish Physiol. Biochem. 5 (1988) 229.
[33] M. Smith, F.B. Jungalwala, J. Lipid Res. 22 (1981) 697.
[34] S. Ramanadham, A. Bohrer, R.W. Gross, J. Turk, Biochemistry 32 (1993) 13499.
[35] A. Üllen, G. Fauler, H. Kofeler, S. Waltl, C. Nusshold, E. Bernhart, H. Reicher, H.J. Leis, A. Wintersperger, E. Malle, W. Sattler, Free Radic. Biol. Med. 49 (2010) 1655.
[36] K.A. Kayganich, R.C. Murphy, Anal. Chem. 64 (1992) 2965.
[37] Y. Sugiura, Y. Konishi, N. Zaima, S. Kajihara, H. Nakanishi, R. Taguchi, M. Setou, J. Lipid Res. 50 (2009) 1776.
[38] F.B. Jungalwala, V. Hayssen, J.M. Pasquini, R.H. Mccluer, J. Lipid Res. 20 (1979) 579.
[39] F.F. Hsu, J. Turk, J. Am. Soc. Mass Spectrom. 11 (2000) 437.
[40] H. Sakai, Y. Tanaka, M. Tanaka, N. Ban, K. Yamada, Y. Matsumura, D. Watanabe, M. Sasaki, T. Kita, N. Inagaki, J. Biol. Chem. 282 (2007) 19692.
[41] Y.C. Ma, H.Y. Kim, Anal. Biochem. 226 (1995) 293.
[42] M. Hermansson, A. Uphoff, R. Kakela, P. Somerharju, Anal. Chem. 77 (2005) 2166.
[43] V. Matyash, G. Liebisch, T.V. Kurzchalia, A. Shevchenko, D. Schwudke, J. Lipid Res. 49 (2008) 1137.
[44] X.N. Li, Y.Z. Liang, F.T. Chau, Chemom. Intell. Lab. Syst. 63 (2002) 139.
[45] S.G. Roussis, R. Proulx, Anal. Chem. 75 (2003) 1470.
[46] J. Meija, J.A. Caruso, Am. Soc. Mass Spectrom. 15 (2004) 654.