# Interpretation of chromatographic and mass spectrometric data from analyses of fatty acid methyl esters

*Application of multivariate methods.*

**Svein Are Mjøs**


Dissertation for the degree philosophiae doctor (PhD)
at the University of Bergen

2006

# Acknowledgements

The work presented in this thesis has been carried out at Norwegian Herring Oil and Meal Industry Research Institute (SSF) and Norwegian Institute of Fisheries and Aquaculture Research (Fiskeriforskning).

Financial support has been provided by:

- Norwegian Research Council (NFR), projects 117234/112 and 136338/140
- International Fishmeal & Oil Manufacturers Association (IFOMA), London, UK
- Denofa, Fredrikstad, Norway
- Margarinfabrikken Norge, Oslo, Norway
- Norsildmel A/L, Bergen, Norway

There are many persons who in some way have influenced and contributed to this work. I am especially grateful to the following:

- My supervisor at the Institute of Chemistry, University of Bergen, Otto Grahl-Nielsen for inspiration, encouragement and guidance.
- Dr. Jan Pettersen, who initiated the work on gas chromatography–mass spectrometry of fatty acids at SSF.
- Sonnich Meier, who is the only person I know that can talk about chromatography and fatty acids for an entire evening at Garage.

All the anonymous and unpaid referees on the papers are gratefully acknowledged for valuable comments on the manuscripts and stimulating questions, and for spending time on an activity that does not give them credit and recognition.

I want to express my gratitude to all my colleagues at SSF/Fiskeriforskning for a positive environment, and to the institute for giving me the opportunity to dig into details about chromatography and mass spectrometry, although the finances in periods have been sparse and the return on investments may be difficult to see.

I would like to thank the people who have provided housing at Finse, where the majority of the papers were written: Inger Margrethe & Helge, Ernst & Nina, Rune & Heidi.

I would also like to thank Ole J. Mjøs for proofreading the thesis.

To all my friends, especially those who go skiing, climbing or kayaking: without all your distractions I may have finished this thesis a bit sooner, but what a boring period it would have been. Your distractions have been greatly appreciated.

Finally, I would like to express my deepest gratitude to my parents, first for their excellent decision to have a second child, and for all love and support throughout my life. Obviously, without them this thesis would not have been possible.

# Summary

The purpose of this work has been to develop tools for fatty acid identification, with special emphasis on poly-unsaturated fatty acids and *trans* fatty acids. The work is based on application of gas chromatography with mass spectrometric detection and the fatty acids are analysed as fatty acid methyl esters (FAME). Various multivariate analytical techniques are applied as tools for interpretation of both chromatographic and spectral information. The ten papers are divided into five subsections with the following topics:

Identification of FAME from shifts in retention indices:

**Papers I** and **II** describe methods for partial identification of fatty acid structure from shifts in retention indices. These methods will basically provide information on the polarity (number of double bonds) and the chain lengths of the fatty acids. The procedure is not restricted to any particular class of fatty acids and may also be used to exclude compounds that are not fatty acids. Two-dimensional fatty acid retention indices (2D-FARI) are introduced in Paper II.

Application to *trans* isomers:

**Papers III** and **IV** are about *trans* isomers of eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA). Paper III provides data about chromatographic properties of the *trans* isomers of EPA and DHA, which can be applied both for identification purposes and for optimisation of chromatographic parameters. A practical application of the data is shown in Paper IV.

Prediction of equivalent chain lengths:

**Papers V** and **VI** describe methods for prediction of equivalent chain lengths (ECL), which is the retention indices commonly applied for fatty acids. Paper V is restricted to polyunsaturated fatty acids with methylene-interrupted double bond systems. Paper VI is not restricted to a particular class of fatty acids, but 2D-FARI data must be available. Thus, the two procedures are complementary methods for ECL predictions.

Identification of FAME from mass spectra:

**Paper VII** and **VIII** are about multivariate analysis of mass spectra of FAME. In Paper VII it is shown that *trans* isomerism in certain positions has significant influence on the mass spectra of polyunsaturated fatty acids. In Paper VIIi it is shown that the number of double bonds in polyunsaturated fatty acids can be determined from selected ions in the mass spectra.

Deconvolution of overlapping chromatographic peaks:

**Papers IX** and **X** are about techniques for deconvolution of overlapping FAME peaks. Transformations for reduction of noise in the mass spectra were introduced in Paper IX. The transformations and the information from Paper VII about differences in the mass spectra caused by *trans* geometry were utilised for deconcvolution of overlapping chromatographic peaks of *trans* isomers in Paper X. Although the focus in this paper is basically on the quantification of the isomers, deconvolution of overlapping peaks has relevance also for the identification of the analytes because spectra and accurate retention times of the compounds are also provided.

# List of publications

**I** S.A. Mjøs: *Identification of fatty acids in gas chromatography by application of different temperature and pressure programs on a single capillary column.* J. Chromatogr. A 1015 (2003) 151–161.

**II** S.A. Mjøs: *Two-dimensional fatty acid retention indices*. J. Chromatogr. A 1061 (2004) 201–209.

**III** S.A. Mjøs: *Properties of trans isomers of eicosapentaenoic acid and eicosahexaenoic acid methyl esters on cyanopropyl stationary phases.* J. Chromatogr. A 1100 (2005) 185–192.

**IV** S.A. Mjøs, S. Meier, O. Grahl-Nielsen: *Geometrical isomerisation of double bonds in acid catalysed preparation of fatty acid methyl esters.* Eur. J. Lipid Sci. Technol. 108 (2006) 315–322.

**V** S.A. Mjøs, O. Grahl-Nielsen: *Prediction of gas chromatographic retention of polyunsaturated fatty acid methyl esters.* J. Chromatogr. A 1110 (2006) 171–180.

**VI** S.A. Mjøs: *Prediction of equivalent chain lengths from two-dimensional fatty acid retention indices.* J. Chromatogr. A 1122 (2006) 249–254.

**VII** S.A. Mjøs, J. Pettersen: *Determination of trans double bonds in polyunsaturated fatty acid methyl esters from their electron impact mass spectra.* Eur. J. Lipid Sci. Technol. 105 (2003) 156–164.

**VIII** S.A. Mjøs: *The prediction of fatty acid structure from selected ions in electron impact mass spectra of fatty acid methyl esters.* Eur. J. Lipid. Sci. Technol. 106 (2004) 550–560.

**IX** S.A. Mjøs: *Spectral transformations for deconvolution methods applied on gas chromatography–mass spectrometry data.* Anal. Chim. Acta 488 (2003) 231–241.

**X** S.A. Mjøs: *Quantification of linolenic acid isomers by gas chromatography–mass spectrometry and deconvolution of overlapping chromatographic peaks.* Eur. J. Lipid Sci. Technol. 106 (2004) 307–318.

# Contents

# Acronyms and symbols

| | |
|---|---|
| 2D-FARI | Two-dimensional fatty acid retention index |
| AMU | Atomic mass units |
| CI | Chemical ionisation (mass spectrometry) |
| CNP | Cyanopropyl (stationary phase) |
| DHA | Docosahexaenoic acid (22:6 *n*–3) |
| DMOX | Dimethyloxazoline (fatty acid derivatives for GC–MS) |
| DUFA | Diunsaturated fatty acid(s) |
| ECL | Equivalent chain length (Equation 6) |
| EFA | Evolving factor analysis |
| EI | Electron impact (ionisation) |
| ENPE | Estimated number of prediction errors |
| EPA | Eicosapentaenoic acid (20:5 *n*–3) |
| FAME | Fatty acid methyl ester(s) |
| FCL | Fractional chain length (Equation 7) |
| FID | Flame ionization detector |
| GC | Gas chromatography / Gas chromatograph |
| *I* | Kovats retention index (Equation 4 and 5) |
| IR | Infra red (spectroscopy) |
| *k* | Retention factor (Equation 1) |
| LC | Liquid chromatography |
| LPG | Latent projective graph |
| *m*/*z* | Mass-to-charge ratio |
| $M^+$ | Molecular ion |
| MI | Methylene interrupted (double bond systems) |
| MI-PUFA | PUFA with methylene interrupted double bonds |
| MLR | Multiple linear regression (Equation 12) |
| MS | Mass specrometry / Mass spectrometer |
| MUFA | Monounsaturated fatty acid(s) |
| NIR | Near infra-red (spectroscopy) |

| | |
|---|---|
| NMI | Non-methylene interrupted (double bond systems) |
| NMI-PUFA | PUFA with non-methylene interrupted double bonds |
| PC | Principal component |
| PCA | Principal component analysis |
| PCR | Principal component regression |
| PEG | Pollyethylene glycol (stationary phase) |
| PLS | Partial least squares |
| PLSR | Partial least squares regression |
| PUFA | Polyunsaturated fatty acid(s) |
| QSRR | Quantitative structure-retention relationship |
| $r$ | Correlation coefficient |
| $r^2$ | Coefficient of determination |
| RMSEC | Root mean squared error of calibration |
| RMSEP | Root mean squared error of prediction (Equation 15) |
| $R_s$ | Chromatographic resolution (Equation 3) |
| SEC | Standard error of calibration |
| SEP | Standard error of prediction (Equation 14) |
| SFA | Saturated fatty acid |
| SIM | Selected ion monitoring (mass spectrometry) |
| TLC | Thin layer chromatography |
| $t_M$ | Hold-up time (elution time for an unretained component) |
| $t_R$ | Retention time |
| $t_R$' | Adjusted retention time ($t_R$–$t_M$) |
| UV | Ultra violet (spectroscopy) |
| $w_b$ | Peak width at baseline (Figure 3) |
| $w_h$ | Peak width at half height (Figure 3) |
| δ | Bias (Equation 13) |

# 1. Background

## 1.1. Fatty acid structure and nomenclature

Fatty acids consist of a carboxylic group connected to a carbon chain (Figure 1). The carbon chain may be saturated or unsaturated, and may contain carbon branches as well as other functional groups. However, the majority of fatty acids in nature have unbranched carbon chains with 4–24 carbons, 0–6 double bonds, and no other functional groups. Fatty acids with odd-numbered carbon chains are present only in minor amounts in most organisms. Minor amounts of fatty acids with carbon chains longer than C24 are also present in marine lipids [1,2].

Several types of fatty acid nomenclature are common, and naming of fatty acids in the literature may vary with what is convenient and with journal policy. Unbranched fatty acids are described by the number of carbons followed by the number of double bonds. Thus, the saturated fatty acid (SFA) in Figure 1a may be denoted ‘C16:0’ or ‘16:0’.

Double bond positions may be described from either end of the molecule. Double bond positions given from the methyl end of the carbon chain are commonly referred to by ‘ω’ or by ‘*n*–’. The monounsaturated fatty acid in Figure 1b may be denoted as ‘20:1 (ω9)’ or ‘20:1 *n*–9’, the latter is usually preferred in chemical literature [3]. Alternatively, the double bond position may be specified by the distance from the carbonyl group as ‘Δ11-20:1’ or as ‘11-20:1’.

Double bonds in polyunsaturated fatty acids (PUFA) are typically separated by a single methylene unit (Figure 1c and d). Double bond systems with this regular pattern are often referred to as *methylene interrupted* (MI) double bonds, and PUFA with this system will be referred to as MI-PUFA in this text. The term *homoallylic* double bonds is equivalent to MI double bonds. In these cases, the complete molecular structure can be described by specifying the number of carbons, the number of double bonds and the position of the double bond system. Thus, the structures given in Figure 1c and d may be described as ‘18:2 *n*–6’ or ‘20:5 *n*–3’. If the ‘ω’ or ‘*n*–’ systems are used for designation of PUFA structure, it is normally taken as granted that all double bonds are methylene-interrupted and have *cis* geometry. In cases where the positions are given from the carbonyl group, all positions are given also for MI-PUFA and the fatty acids in Figure 1c and d are named 9,11-18:2 and 5,8,11,14,17-20:5.

Double bond systems that do not have the regular methylene interrupted patters may have conjugated (Figure 1e) or isolated (Figure 1f and g) double bonds. The latter

are often referred to as non-methylene interrupted (NMI) systems. In these cases, it is common to specify the distance of all double bonds from the carbonyl group.

Reference can also be made to specific double bond positions or groups of isomers, *e.g.* '*n*–9 double bonds', with no assumptions about position or geometry of the rest of the double bonds in the molecule. Thus, the isomer shown in Figure 1f has an *n*–6 double bond, even though it is separated from the next bond by more than one methylene group.

The terms '*cis*' and '*trans*' are commonly applied to describe the geometries of double bonds in fatty acids instead of '*Z*' (zusammen) and '*E*' (entgegen), which is more common in general organic chemistry. The geometries are often described by a single letter '*c*' or '*t*', which is combined with the double bond position. The NMI fatty acid in Figure 1g can be referred to as 7*c*,11*t*-18:2, and the isomer of 18:3 *n*–3 with *trans* geometries in the Δ9 and Δ12 position (Figure 1h) is 9*t*,12*t*,15*c*-18:3. If the geometries are not specified, the double bonds are usually expected to have *cis* geometry.

Fatty acids are sometimes denoted by their systematic names, *e.g.* 20:4 *n*–6 is referred to as 5,8,11,14-eicosatetraenoic acid and 7*c*,11*t*-18:2 may be referred to as (*Z*,*E*)-7,11-octadienoic acid. Common names, *e.g.* Arachidonic acid (20:4 *n*–6), or abbreviations for systematic names, *e.g.* EPA for eicosapentaenoic acid (20:5 *n*–3) or DHA for docosahexaenoic acid (22:6 *n*–3), are also applied.

Even though the structures described above cover the majority of common fatty acids, there are a large number of less common fatty acids with various structures. The carbon chain may contain triple bonds, branches, as well as saturated and unsaturated carbon rings [4,5]. Oxygen may be introduced in the carbon chain in form of hydroxy groups, oxo groups, furan rings or additional carboxyl groups [4, 5]. Other heteroatoms may also be present, *e.g.* halogens [6,7].

The majority of fatty acids are esterified to larger lipid molecules; only small amounts are present in free form in living organisms as well as food matrices. Lipid molecules are traditionally classified into neutral and polar lipids. Common neutral lipids are triacylglycerols, used as energy reserve in most organisms; wax esters, the energy reserve in certain marine species; cholesteryl esters and free fatty acids. In most tissues, the majority of polar lipids are phospholipids from cell membranes.

Fatty acids are mainly analysed by gas chromatography as their corresponding fatty acid methyl esters (FAME). The preparation of FAME involves extraction of the lipid molecules from the sample matrix, breaking of the ester bonds, and formation of methyl esters. The two last steps may be combined by trans-esterifying the lipids directly with acid or base in methanolic solution. The process of converting free or esterified fatty acids into methyl esters are commonly referred to as methanolysis, (trans-) esterification, or methylation. Base-catalysed methods, typically carried out

by methanolic NaOH, will not methylate free fatty acids and require anhydrous conditions. Several acids are used as catalysts for one-step methylation, including $H_2SO_4$, HCl and $BF_3$. There are positive and negative aspects with all these catalysts. Other fatty acid derivatives than FAME are applied for special purposes. An extensive review over fatty acid derivatisation is given in [8].

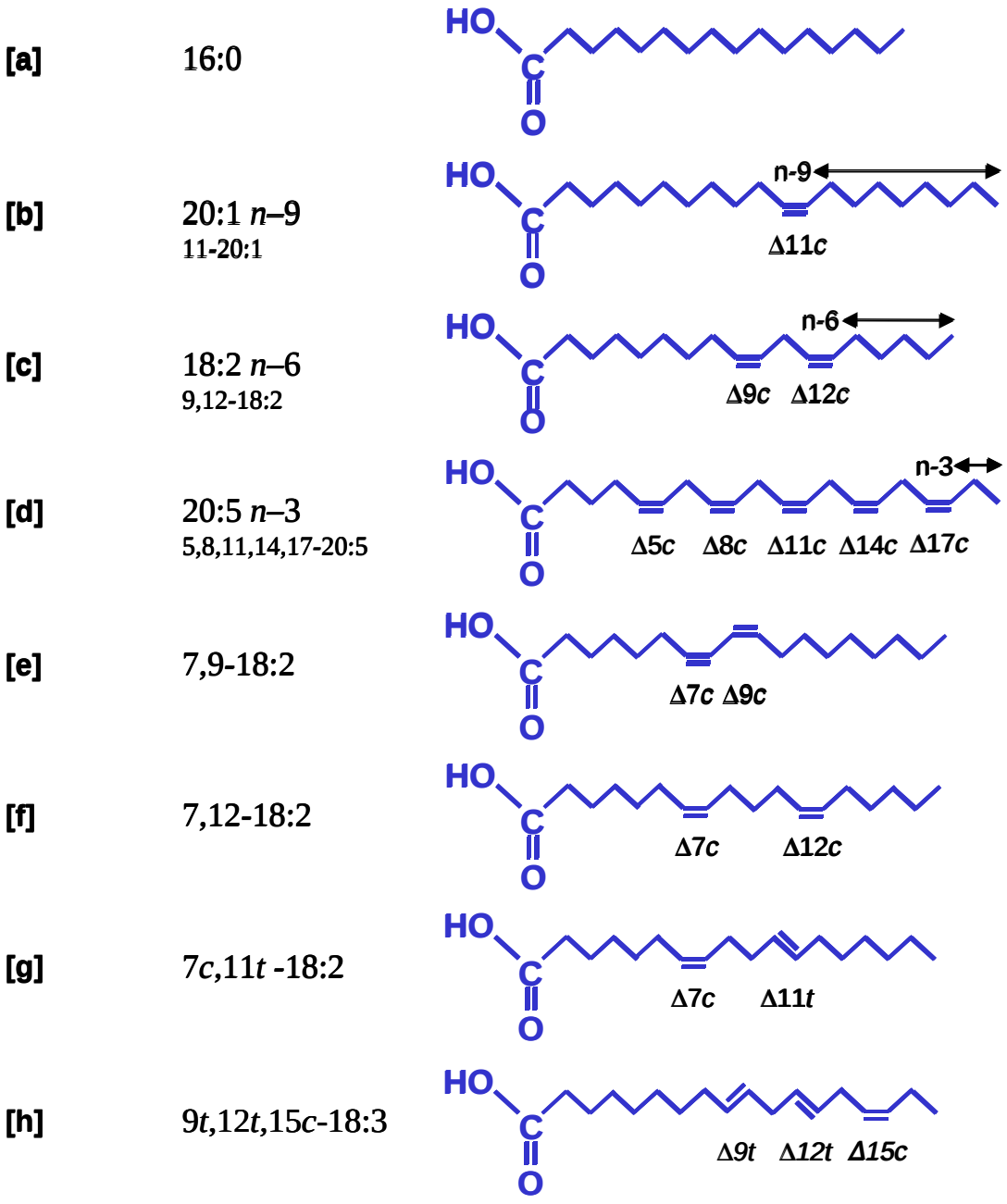



**Figure 1** *Overview over fatty acid structure and nomenclature.*

# 1.2. Gas Chromatography

## *1.2.1 Principles of separation*

Modern gas chromatography (GC) is typically based on open tubular capillary columns varying in length from 10–100 m and with internal diameters from 0.1–0.5 mm. The principle of separation in open tubular gas chromatography is explained in Figure 2. The retention of a compound is determined by its distribution between the stationary phase and the mobile phase. The distribution can be expressed by the retention factor, $k$ [9]:

$$k = \frac{\text{amount of solute in stationary phase}}{\text{amount of solute in mobile phase}} \qquad (1)$$

$k$ depends on the solubility of the solute in the stationary phase ($k$ increases with increased solubility), the thickness of the stationary phase ($k$ increases with increased thickness), column diameter ($k$ decreases with increased diameter), and the temperature ($k$ decreases with increased temperature).

As long as $k$ is constant throughout the separation process, it is related to the retention time of the compound by the following equation:

$$k = \frac{t_R'}{t_M} = \frac{t_R - t_M}{t_M} \qquad (2)$$

where $t_R$ is the elution time of the solute, $t_M$ is the elution time of an unretained component (hold-up time) and $t_R$' is the adjusted retention time ($t_R - t_M$). $k$ will be constant only as long as the temperature is constant. Equation 2 is therefore not valid for temperature-programmed gas chromatography.

Retention as described in Figure 2 and by the equations above is an idealised model, where it is assumed that solutes in the stationary phase behave as ideal solutions. Deviations from ideal conditions may be caused by surface effects between the stationary phase and the carrier gas, uneven distribution and composition of the stationary phase, adsorption of the analytes, displacement effects and interactions between analytes. Deviations from ideal conditions may be especially large for analytes with low volatility and low solubility in the stationary phase or when the capacity of the stationary phase is overloaded.

It should also be emphasized that the proportions shown in Figure 2 do not correspond with dimensions in real capillary columns. In modern columns the internal diameter is typically 1000 times larger than the thickness of the stationary phase. The solutes also elute in broad bands.

The term ‘carrier gas velocity’ in GC normally refers to the average gas velocity in the column, which is the length of the column divided by the elution time for an unretained component. Because of the high compressibility of the carrier gas and large pressure drop in the column the actual carrier gas velocity is higher in the end than at the head of the column.

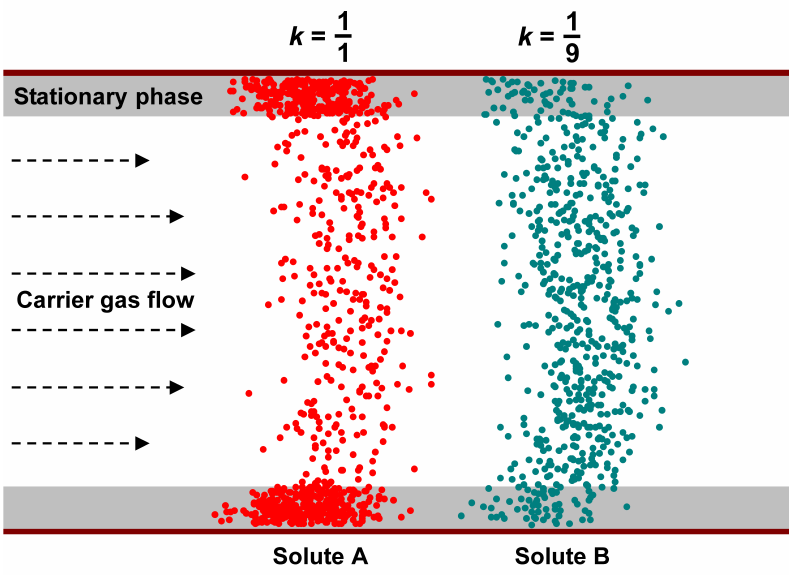



**Figure 2** *Principle of separation in open tubular gas chromatography. The two solutes A and B are separated because they differ in retention factor, k. Solute A is equally distributed between the stationary phase and the mobile phase (k = 1). Solute B moves faster through the column because a larger fraction of the molecules (90%) is in the mobile phase (k = 0.11).*

Various parameters related to chromatographic separation are based on the assumption that the shapes of chromatographic peaks resemble normal distribution curves as shown in Figure 3 [9]. However, non-ideal conditions, like adsorption of the analytes to active sites, dead volumes, or incomplete solvation of analytes will lead to deviations from a normal distribution, and various alternative peak shape models have been proposed [10,11]. Several parameters are used for describing asymmetry of peaks [12,13].

The separation beween two chromatographic peaks, A and B, may be described by the peak resolution, $R_s$, which is defined as:

$$R_s = \frac{2\,(t_{R(A)} - t_{R(B)})}{w_{b(A)} + w_{b(B)}} = \frac{2\,\Delta t_R}{w_{b(A)} + w_{b(B)}} \qquad (3)$$

where $t_R$ is retention time and $w_b$ is the peak width at baseline. The width at baseline may be difficult to estimate and the peak width at the half height of the peak, $w_h$, is therefore often used as an alternative to $w_b$ (Figure 3).

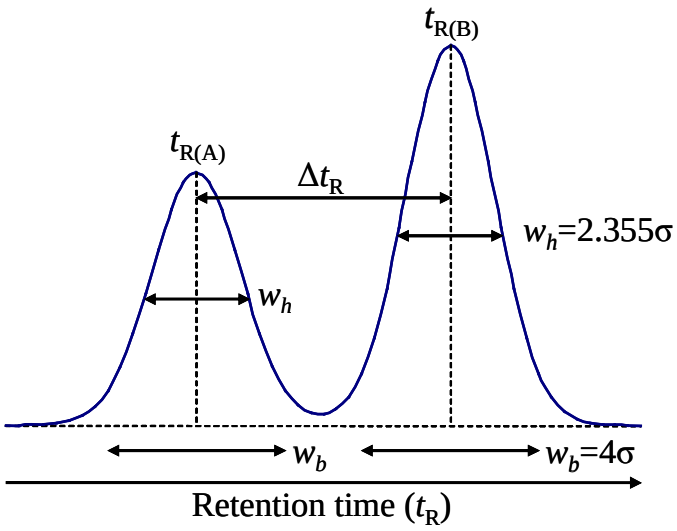



**Figure 3** *Two simulated chromatographic peaks (A and B) with retention times $t_{R(A)}$ and $t_{R(B)}$. σ is the standard deviation of the normal distribution curves. The peak width at baseline is defined as 4σ [9]. The peak width at half height is 2.355σ.*

### *1.2.2. Stationary phases in gas chromatography of fatty acids*

Since the early works of Martin and James on gas chromatographic separation of fatty acids with a silicone coated stationary phase [14], a wide range of phases have been applied for the separation of free and derivatised fatty acids. The development of GC and stationary phase technology for the separation of fatty acids has recently been reviewed by Ackman [15]. Today, polar columns are usually preferred for separation of complex fatty acid mixtures, especially when highly unsaturated fatty acids or *trans* fatty acids are analysed.

There are two types of polar phases that dominate: the polyethylene glycol (PEG) columns and the cyanopropyl (CNP) columns. The polar functional groups in these phases are hydroxy (OH) and cyano (CN) groups. The CNP phases consist of a siloxane polymer with cyanopropyl substituents. The cyanopropyl groups are often combined with less polar groups, *e.g.* methyl or phenyl, and CNP columns are therefore available in a large range of polarities. PEG is usually not mixed with other groups and common PEG columns show little variation in polarity.

This work is mainly based on the application of cyanopropyl columns, a PEG column was applied in parts of Paper IV and an apolar column (100% methyl polysiloxane) was applied in parts of Paper X. Two CPN phases have been applied

in this project. SP-2560 (Supelco, Bellefonte, PA, USA) is a highly polar 100 % CNP column. BPX-70 (SGE, Ringwood, Australia) is less polar than SP-2560 and has aromatic groups introduced in the polymer backbone.

A typical elution pattern of a reference mixture with saturated and *cis* unsaturated FAME on BPX-70 is shown in Figure 4. In general, FAMEs with the same number of carbon atoms elute according to the number of double bonds. However, because *n*–3 PUFA are retained more strongly than *n*–6 PUFA, 20:4 *n*–6 elutes before 20:3 *n*–3. As shown for C20 and C22, there is a substantial overlap between the different chain lengths. In real samples of marine origin there will be corresponding broad regions of C16 and C18 because 16:4, 18:4 and 18:5 can be present. Because of the overlap in chain lengths, the identification of fatty acids from retention times is a difficult task. The picture is further complicated by the presence of odd-chain PUFA, *e.g.* 21:5 *n*–3, which is present in marine lipids [15,16].

A characteristic feature of CNP phases is that the polarity shows a strong dependence on temperature. This attribute is a central issue in Papers I–III. At low temperature, the polarity of the BPX-70 phase is lower than the polarity of PEG columns [17]. While the properties of PEG columns are nearly unaffected by the temperature, the polarity of BPX-70 increases linearly with increasing temperatures [17] and BPX-70 is considerably more polar than PEG columns at the temperatures typically applied for analyses of FAME (>140 °C). The effects of varying chromatographic conditions on the elution pattern are shown in Paper II.

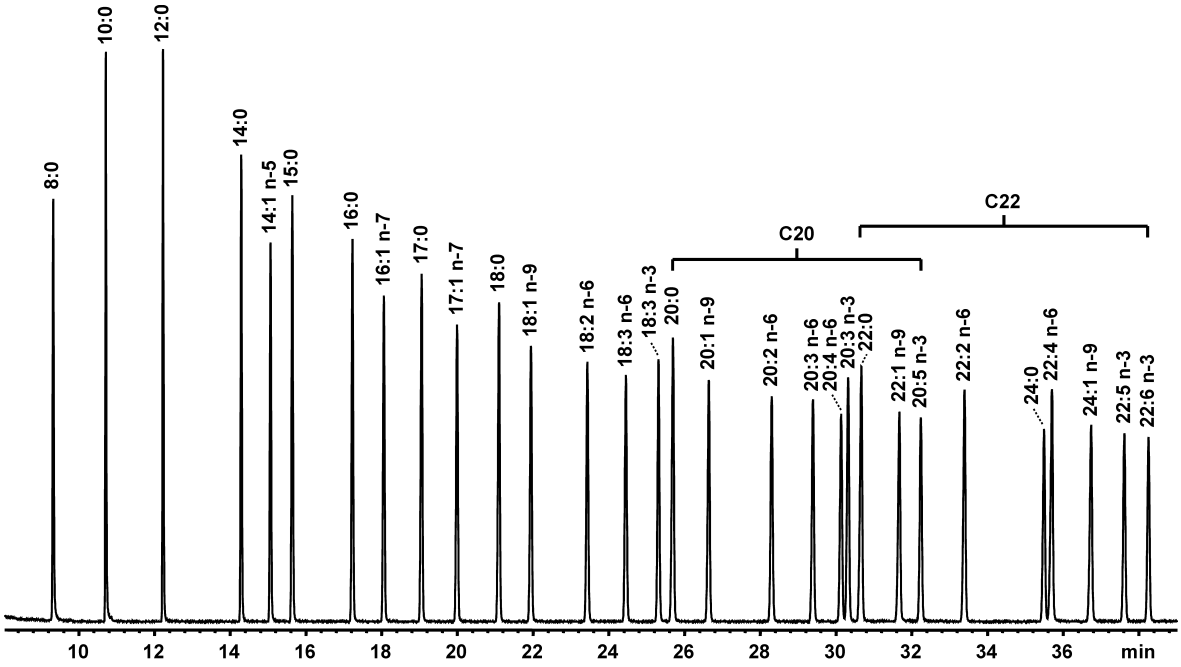



**Figure 4** *Elution pattern of a FAME reference mixture (GLC-461, Nu-Chek Prep, Elysian, MN, USA) on BPX-70. The chromatographic parameters are as described for program 1 in Paper II.*

### 1.2.3. Retention indices and ECL values

On a retention index scale, the chromatographic retention of a compound is described relative to the retention of a series of homologues. The most common system is the Kovats indices [18], where *n*-alkanes are applied as reference compounds. At isothermal conditions there is a linear relationship between log $t_R$' and the number of carbon atoms in homologous series, and the retention index, *I*, for a compound, *x*, can be calculated by:

$$I_{(x)} = 100n \frac{\log t'_{R(x)} - \log t'_{R(z)}}{\log t'_{R(z+n)} - \log t'_{R(z)}} + 100z \qquad (4)$$

where $t_R$' is adjusted retention times of the compound of interest and two *n*-alkanes eluting on each side of the compound, *z* represents the number of carbon atoms in the *n*-alkane eluting before *x*, and *n* is the difference in carbon atoms between the two *n*-alkane references. For maximal accuracy, it is recommended that *n* is one. Kovats indices acquired at isothermal conditions are assumed to be invariant to differences in column dimensions and carrier gas flow, but are highly dependent on the stationary phase and also influenced by the oven temperature. Thus, *I* acquired at a certain stationary phase at a certain temperature is a characteristic property for a compound that can be used for identification purposes.

The use of retention indices has been extended to temperature-programmed gas chromatography, where there exists an approximately linear relationship between retention times and the number of carbon atoms in a homologous series. In temperature programmed GC, *I* is generally calculated by the van den Dool and Kratz formula [19]:

$$I_{(x)} = 100n \frac{t_{R(x)} - t_{R(z)}}{t_{R(z+n)} - t_{R(z)}} + 100z \qquad (5)$$

The parameters *n*, *x* and *z* are the same as in Equation 4. Equation 5 gives the same results whether applied with gross retention times or adjusted retention times. As for Equation 4, it is recommended that *n* is one. In addition to application of the equations above, various approaches based on higher order regressions and other approaches are applied [20–23, Paper I–II].

Although isothermal retention indices are assumed to be independent of carrier gas flow, column dimensions and phase ratios, temperature programmed indices are not. Increased temperature gradient, column length or phase ratio, or decreased carrier gas flow rate, will move *I* in the same direction as increased temperature in isothermal chromatography [24].

Although Kovats indices are the dominating general-purpose retention index system, a large number of alternative series with other calibration compounds than *n*-alkanes

have been applied for special purposes. The motivations for using other calibration standards than alkanes are basically the following:

- *n*-alkanes cannot be detected by several common detection methods such as negative ion chemical ionisation mass spectrometry, electron capture detectors and element specific detectors.

- Retention indices based on molecules with the same functional groups as the analytes of interest are often more reproducible, and vary less with chromatographic conditions than indices based on *n*-alkanes [25].

- A second calibration mixture of *n*-alkanes is not necessary if the retention index scale is defined by some of the analytes of interest.

- *n*-alkanes have poor chromatographic properties on highly polar stationary phases.

Alternative retention indices for various purposes have been extensively reviewed elsewhere [26,27]. The most successful approach may be the use of equivalent chain lengths (ECL) [28] for fatty acid analysis. The ECL system is based on saturated unbranched FAMEs as reference compounds and ECL values for the references are by definition equal to the number of carbons in the alkyl chain. Thus, ECL for compound *x* at isothermal conditions can be calculated by:

$$\mathrm{ECL}_{(x)} = n \frac{\log t'_{\mathrm{R}(x)} - \log t'_{\mathrm{R}(z)}}{\log t'_{\mathrm{R}(z+n)} - \log t'_{\mathrm{R}(z)}} + z \qquad (6)$$

where $t_{\mathrm{R}}$' is adjusted retention times of the compound of interest, *x*, and two saturated FAMEs eluting on each side of the compound. *z* represents the number of carbon atoms in the carbon chain of the saturated FAME eluting before *x*, and *n* is the difference in the number of carbon atoms between the two references. ECL values at temperature-programmed GC can be calculated by the same methods as *I*.

The fractional chain length is defined as the difference between the ECL value and the number of carbons in the fatty acid chain of the FAME molecule and is calculated by:

$$\mathrm{FCL}_{(x)} = \mathrm{ECL}_{(x)} - \mathrm{NC}_{(x)} \qquad (7)$$

where NC is the number of carbons in the fatty acid chain. It follows from the definition of ECL that FCL of the saturated unbranched FAMEs are zero. The unsaturated FAMEs, which on polar columns elute after the saturated FAME with the same number of carbons, have positive FCL values and FCL is used as an indication of the polarity of a fatty acid.

### *1.2.4. Identification of fatty acids from retention times and ECL-values*

Various strategies have been applied for the prediction of chromatographic properties of fatty acids that are not available as reference compounds. A much used strategy has been to assume that members in the same homologous series will have similar behaviour relative to the saturated analogues, *i.e.* members in a homologous series have almost identical FCL values. These relationships have been widely applied in isothermal chromatography, where fatty acids are identified from parallel lines drawn between members of the same homologous series in plots of ECL values or log $t_R$' against the number of carbons in the molecule [29–31]. The assumption of similar FCL values for members in the same homologous series can be expected to be accurate as long as interactions between the carbonyl group and the double bond system can be neglected, but may give inaccurate predictions for molecules with double bonds close to the carbonyl group.

Accurate prediction of ECL values is more challenging with temperature-programmed chromatography than with isothermal chromatography, particularly with CNP phases where the assumption of constant FCL within a homologous series is inaccurate. Because the heaviest members in the series elute at higher temperature than lighter homologues they elute from a column that appears to be more polar, and the FCL values may therefore increase with chain length within the series.

Another strategy for prediction of ECL values has been to assume that the influence of double bonds are additive, and that FCL values of PUFAs can be predicted by summing the FCL values of monounsaturated fatty acids with double bonds in the corresponding positions [32–35] or by adding FCLs for monoenes to FCLs of other PUFAs [34,35], *e.g.* FCL for 18:3 *n*–3 is predicted by summing FCL for 18:2 *n*–6 and FCL for 18:1 *n*–3. The accuracies of these calculations are low because MI double bonds behave differently than the sum of the corresponding isolated double bonds, and additional correction factors must be introduced [32–35]. The availability of relevant FCL data for the monoenes are also limited. Alternative methods for prediction of ECL values are proposed in Papers V and VI.

## 1.3. Mass spectrometry of fatty acids

### *1.3.1. Principles of mass spectrometry*

In mass spectrometry (MS) with electron impact (EI) ionisation the molecules in gas phase are bombarded with high-energy electrons and form radical cations. Unstable radical cations will decompose in the mass spectrometer and the degree of fragmentation depends on how well the molecule can stabilise the positive charge. The resulting fragments are separated according to their mass-to-charge ratio (*m*/*z*). The charge of the detected fragments can be assumed to be +1, and their masses will

therefore be known. The molecular weight can be told from the molecular ion ($M^+$), and the fragmentation pattern can give important information about functional groups and isomerism.

### *1.3.2. Mass spectrometry of fatty acid methyl esters*

The number of carbons and number of double bonds can easily be determined from high quality mass spectra of FAME, at least for analytes with zero to three MI double bonds. The molecular ions are usually visible, and the distributions of fragments with low masses show characteristic patterns. Examples of mass spectra of common C18 FAMEs are given in Figure 5a to e. The most important ions are listed in Paper VIII and a brief summary is given below.

In normal (unbranched) saturated fatty acids, fragmentation is dominated by the McLafferty rearrangement giving the base peak at *m/z* 74. Abundant peaks at *m/z* 87, 143, 199 and 255 arise from loss of the neutral aliphatic radicals with the general formula $[(CH_2)_nCOOCH_3]^+$ [36]. The molecular ion (*m/z* 298 in Figure 5a) is usually abundant.

Double bonds in linear alkenes and unsaturated fatty acid methyl esters tend to migrate in the molecular ion prior to fragmentation, making the determination of double bond position in unsaturated fatty acids uncertain. There are therefore no ions that serve to indicate the position or the stereochemistry of the double bond in monoenes [36,37]. The spectrum of 18:1 *n*–9 is shown in Figure 5b. The fragmentation pattern in monoenes is dominated by a series of ions with the formula $[C_nH_{2n-1}]^+$ (*m/z* 55, 69, 83, and 97). The Molecular ion is seen at *m/z* 296. The more abundant peaks at 264 and 265, arising from loss of methanol and methoxy radical, may also serve as indicators for molecular weight in low quality spectra [37].

The spectrum of 18:2 *n*–6 is shown in Figure 5c. In MI dienes a series of ions $[C_nH_{2n-3}]^+$ (*m/z* 67, 81, 95, 109) dominate the spectrum at low masses. The $[C_nH_{2n-1}]^+$ series is also abundant at masses 55 and 69. The molecular ion (*m/z* 294) is more abundant than in monoenes. Loss of methoxy radical is also seen at *m/z* 263.

The spectrum of 18:3 *n*–6 is shown in Figure 5d. In MI-PUFA the series $[C_nH_{2n-5}]^+$ dominates the pattern with abundant fragments at *m/z* 79, 93, 107 and 121. The $[C_nH_{2n-3}]^+$ series is also abundant (*m/z* 67, 81, 95 and 109). Fragments from the $[C_nH_{2n-1}]^+$ series are seen at *m/z* 55 and 69. In MI-PUFA (not including dienes) there are rules for predicting the position of the double bond system from diagnostic ions. The position of the first double bond counted from the methyl end of the carbon chain can be determined from the ions $[C_{n+5}H_{2n+8}]^{+\bullet}$ where n is the number of carbons from the methyl end to the first double bond. Thus, *n*–3, *n*–6 and *n*–9 families will have abundant ions of *m/z* 108, 150 and 192 respectively [38–40].

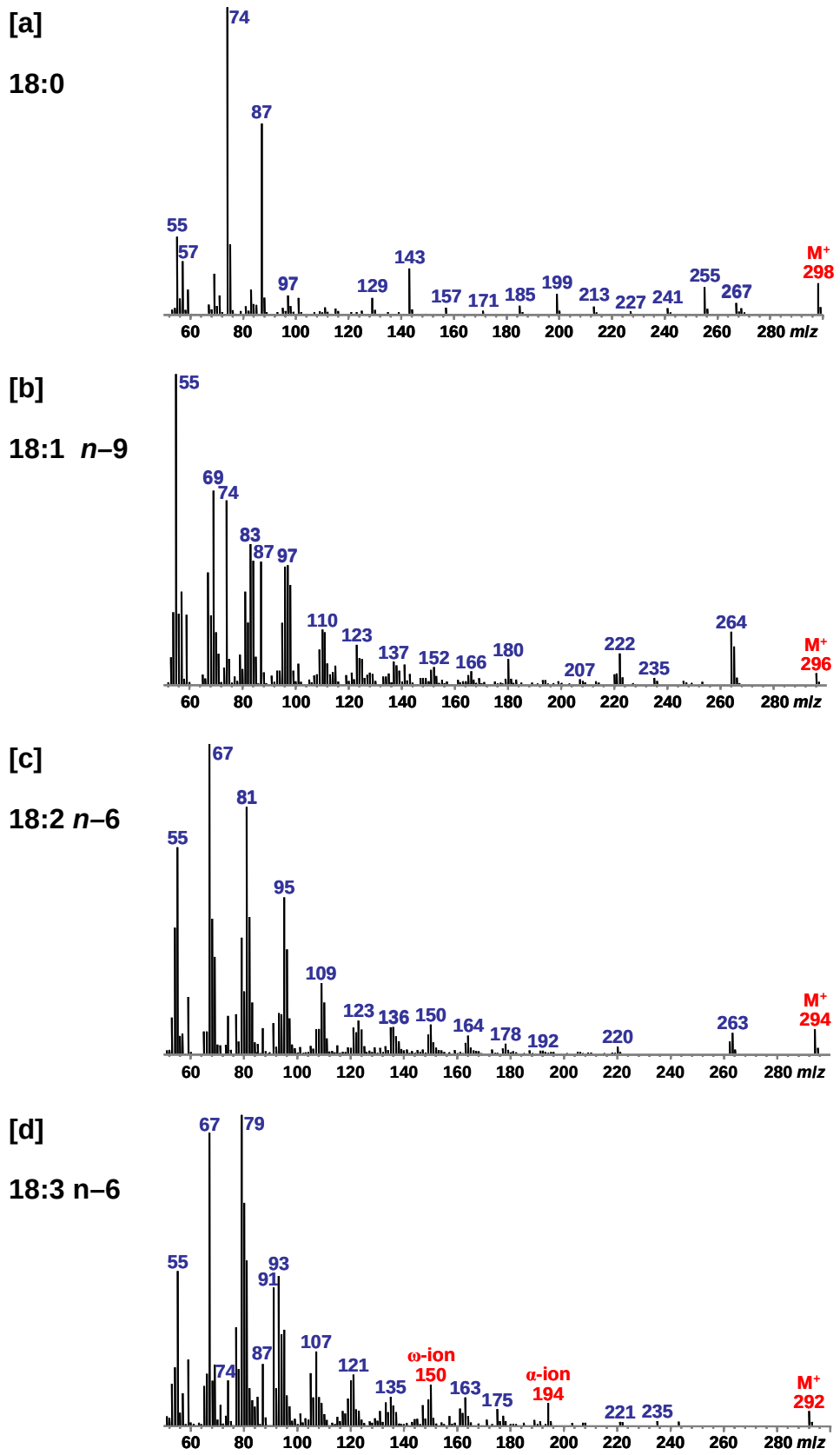



**Figure 5** *Electron impact mass spectra of selected C18 fatty acid methyl esters. (See also Figure 21)*

It has also been reported that the position of the first double bond counted from the carbonyl group can be determined from the ions $[C_{n+6}O_2H_{2n+6}]^{+\bullet}$ where n is the distance from the carbonyl group to the first double bond (Δ-position) [38,39]. MI-PUFA with the first double bonds in the Δ6 or Δ9 positions would have abundant ions of *m/z* 194 and 236 respectively. The diagnostic ions used for determination of the double bond positions relative to the carbonyl group and to the methyl end of the carbon chain will be referred to as α- and ω-ions respectively. Their abundances in electron impact (EI) mass spectra are often low, but can be considerably enhanced with reduction of the ionization energy [39].

Mass spectra of several NMI trienes are presented in [41]. These spectra show different patterns in the lower mass region than MI-PUFA. *m/z* 79, usually the base peak in MI-PUFA, is suppressed and both *m/z* 67 and 81 are of higher abundance.

### *1.3.3. Alternative derivatives and ionization modes*

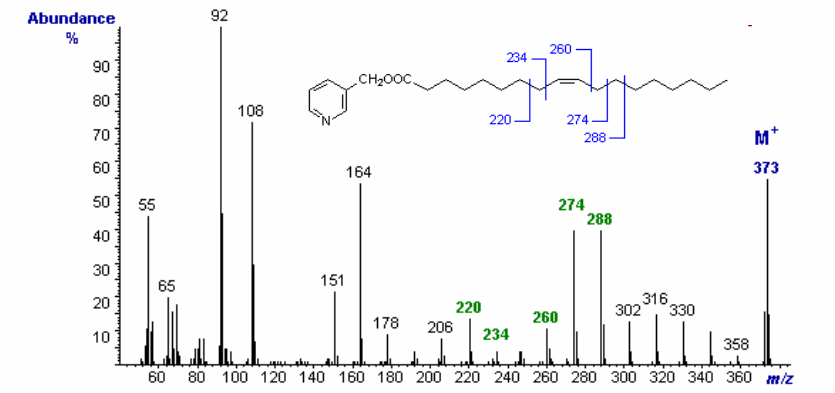
Although FAME give some information about fatty acid structure, more informative spectra about positional isomerism are often achieved using other derivatives. By introducing a nitrogen containing ring in the molecule, the molecular ion is stabilised and double bond migration is reduced. Pyrrolidide, picolinyl and dimethyl oxazoline (DMOX) derivatives are especially useful for determination double bond positions in fatty acids [42–46].

The spectrum of the 18:1 *n*–9 picolinyl ester is shown in Figure 6a. Fragmentation of the carbon chain in the fatty acids gives rise to a series of abundant ions spaced by 14 atomic mass units (AMU). The position of the double bond can be determined from the gap of 26 AMU between *m/z* 234 and 260. The abundant ions with *m/z* 220, 274 and 288 are also of diagnostic importance.

Similar patterns can be applied for the determination of double bond positions also in more unsaturated fatty acids. However, the signals from diagnostic ions are usually weaker and the number of interfering ions from other fragments increase with the number of double bonds. This can be seen in the spectra of 18:3 *n*–3, 20:5 *n*–3 and 22:6 *n*–3 in Figure 6b–d. Thus, interpretation of spectra from highly unsaturated fatty acids is challenging. Since unknown fatty acids are rarely among the most abundant in a lipid sample it is often difficult to acquire pure spectra of sufficient quality.

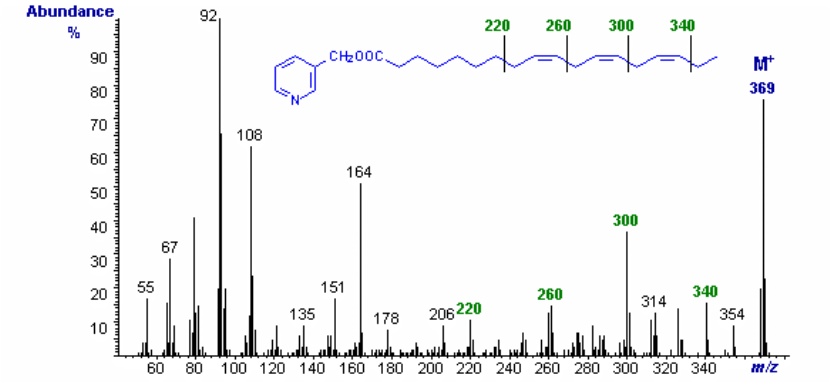
**[a]**

**18:1 *n*–9**



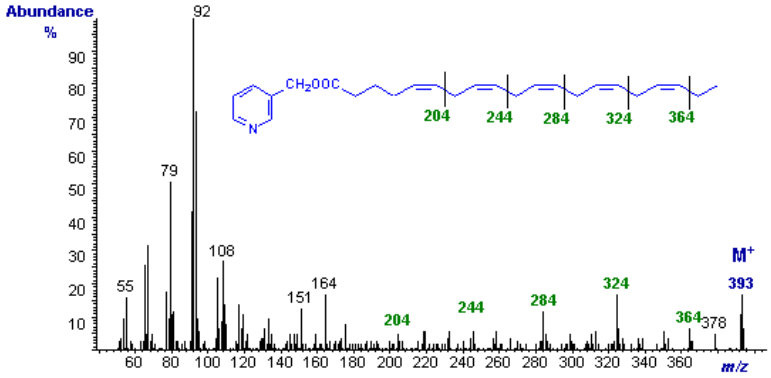

**[b]**

**18:3 *n*–3**




**[c]**

**20:5 *n*–3**




**[d]**

**22:6 *n*–3**

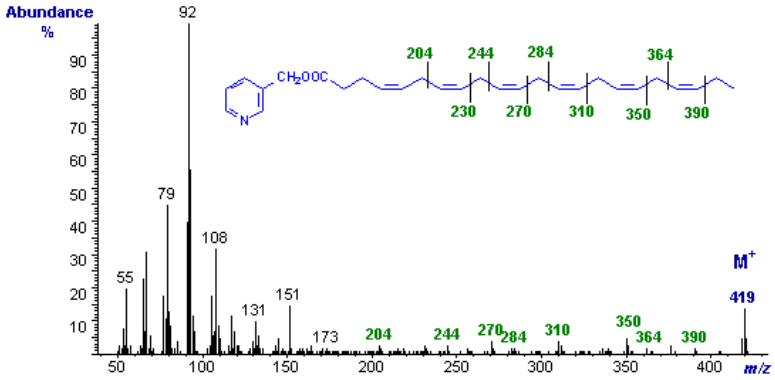



**Figure 6** *Electron impact mass spectra of selected fatty acid picolinyl esters reproduced with permission from [41].*

More distinctive fragmentation patterns in EI-MS can often be obtained by derivatisation of the double bonds. Common procedures are deuteration, silylation [47], or preparation of dimethyl disulfide adducts [48–51]. However, many of these methods are not suitable for complex mixtures because of incomplete reactions, side reactions or decomposition of derivatives at high temperatures. Some of the methods substantially increase the weight of the isomers, which may cause problems with analysis of C20 and C22 fatty acids by GC. Thorough reviews of these and other methods for determination of double bond positions are given elsewhere [8, 52–54].

An alternative to electron impact mass spectrometry is to use chemical ionization (CI). In CI-MS, the charge is transferred to the analytes from an ionized reactant gas. Common reactant gases are methane [55,56], isobutane [57] ammonia [57] and acetonitrile [39,58–60]. Protons are abstracted from the reagent gas leading to abundant $[M+1]^+$ ions. In some cases the entire reactant molecule is abstracted. Because the ionisation energy is lower, there is less fragmentation with CI-MS than with EI-MS and the molecular weight can usually be determined from the $[M+1]^+$ ions [55,56] or higher masses if larger parts of the reactant ions are absorbed.

Because there is less fragmentation, there are normally no diagnostic ions that may provide information about the double bond positions with methane as reagent gas [56]. However, it has been shown that the α and ω ions used for determination of the double bond position in PUFA are abundant in CI spectra of FAME when acetonitrile is applied as reactant gas [39]. It has also been shown that acetonitrile CI-MS spectra of FAME contain diagnostic ions for the positions and geometries of double bonds in conjugated linoleic acid [59].

### *1.3.4. Visual and multivariate interpretation of mass spectra*

Mass spectra of unknown compounds are usually interpreted by visual inspection of the presence or absence of certain ions, *e.g.* the diagnostic ions for double bond positions in Figures 5 and 6. This method has certain drawbacks. If the diagnostic ions are weak, spectra of high quality are required. There are also many cases where there exist no diagnostic ions, *e.g.* the position of double bonds in monoenoic FAME or the geometry of double bonds cannot be determined from the presence or absence of ions. There is also a human factor involved, visual interpretation of mass spectra requires a skilled interpreter and the process is often time consuming.

An alternative strategy is to analyse the spectra with mathematical methods that better utilize small differences in the relative proportions of the masses. Multivariate analysis of spectra is essential in interpretation of near-infrared (NIR) spectra [61] and has also found widespread use with infra-red (IR) [62], ultra-violet (UV) [63], nuclear magnetic resonance [61] and fluorescence spectra [61, 64].

The use of multivariate interpretation is more problematic with mass spectra than with optical methods. While the signals from a certain functional group will appear at (or near) the same wavelengths in for example IR, NIR and UV spectra, the signals in mass spectra are often from loss of functional groups. Thus, signals that can be associated with the functional group may appear at different *m/z*, depending on the structure of the rest of the molecule. In addition, the fragmentation in mass spectrometry is often dominated by fragmentation of the carbon skeleton of the molecule, and small differences in the carbon connectivity may give rise to large differences in the mass spectra. To a certain degree, these problems can be solved by transformation of the spectra to new *spectral features* [65–69].

In molecules with long straight-chain carbon skeletons, such as most FAME, The EI spectra are dominated by a series of low-mass ions that are always present. Ions in this region are suitable for multivariate interpretation, and EI spectra of FAME and similar molecules can be analysed without prior transformations. The double bond positions in monoenoic fatty acid ethyl esters can for instance be determined from multivariate regression on the spectra, even though there are no diagnostic ions (in the classical sense) for the positions [70]. Similar approaches have also been used for the determination of double bond positions in monounsaturated acetates [71,72] and long chain alcohols [71]. Another example is the identification of *trans* geometry in unsaturated fatty acids [73]. An advantage with the statistical approach is that the reliability of the identification can be calculated from the error estimates of the regression or classification. Classic visual mass spectral interpretation provides no such information.

## 1.4. Multivariate methods

Various forms of multivariate mathematical methods, such as principal component analysis (PCA) or partial least squares regression (PLSR), are applied in Papers I–X. A brief introduction to these methods is given below.

### *1.4.1. The nature and representation of multivariate data*

In datasets with a large number of variables it is often difficult to achieve a good picture of the nature of the data and the correlation between the variables. One or two variables can be illustrated on a single surface (*e.g.* a paper or screen), typically in the form of a one-dimensional bar plot or a two-dimensional xy-scatterplot. A third variable can be plotted in three-dimensional xyz-scatterplots, which is a projection of the three dimensions onto a two-dimensional surface.

When the number of variables increase beyond three, most people will have problems perceiving the structure of the data. A number of methods exist that aim at reducing the dimensionality in systems with many variables. Most methods use the

correlation between the original variables to reduce the dimensionality. The original multi-dimensional space may be projected onto so-called *latent variables* [69,74].

### *1.4.2. Matrices and vectors*

Data are often given in matrices where the values for several *objects* (or cases) are described by several measured variables. Such a matrix is illustrated in Figure 7, where eight variables have been measured for 10 objects. Each number in the matrix is referred to as an *element*, each variable is represented by a *column vector* and each sample by a *row vector*.

Column vectors and row vectors can be multiplied; the product is a matrix with dimensions corresponding to the number of elements in the column vector and the row vector. The multiplication of two vectors, $\mathbf{tp}^T$, giving the *outer-product* is illustrated in Figure 8. There is also an *inner-product* of the two vectors that is a scalar found by $\mathbf{p}^T\mathbf{t}$. Brief reviews of common matrix and vector computations used in chemometrics are given in [69,75].

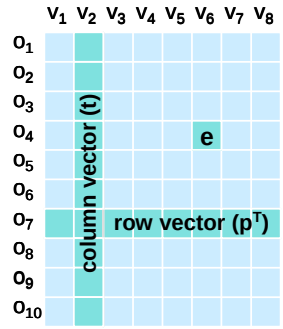



**Figure 7** *10 x 8 matrix,* **X***, with column vector* **t***, row vector,* $\mathbf{p}^T$ *and element,* e*. By convention, vectors are column vectors, the superscript,* $^T$*, on* $\mathbf{p}^T$ *means transposed and denotes that* $\mathbf{p}^T$ *is a row vector.*

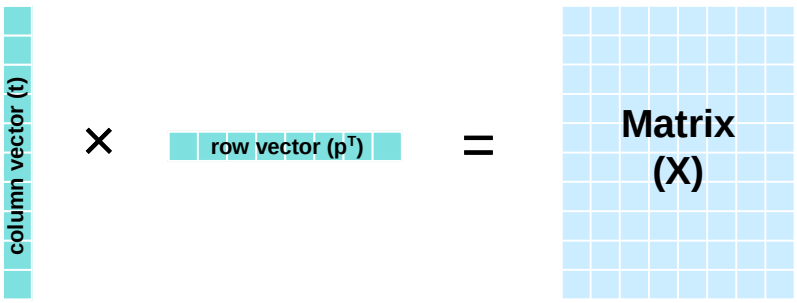



**Figure 8** *Vector multiplication,* $\mathbf{tp}^T$*, giving the outer-product,* **X***.*

### *1.4.3. Principal component analysis (PCA)*

Principal component analysis (PCA) is a technique for extracting structure from high-dimensional datasets. The original variables are projected to latent variables, which in PCA is called *principal components* (PC). The principle of PCA is briefly described below; a comprehensive tutorial on PCA is given in [76].

As illustrated in Figure 8, multiplication of a column vector and a score vector gives a matrix. Matrices can therefore be explained by sets of column and row vectors. When extracting the first principal component the goal is to find the column vector and the row vector that gives the best representation of the variation in the original matrix, **M**. These two vectors, which is called *score vector* and *loading vector*, is the first principal component, $\mathbf{C}_1$ (Figure 9).

With real analytical data, $\mathbf{C}_1$ will never give a perfect description of **M** and there will be a residual matrix $\mathbf{E}_1$, which is found by subtracting each element in $\mathbf{C}_1$ from the corresponding element in **M**:

$$\mathbf{E}_1 = \mathbf{M} - \mathbf{C}_1 = \mathbf{M} - \mathbf{t}\mathbf{p}^T \qquad (8)$$

Another way to describe the extraction of $\mathbf{C}_1$ is that the algorithm seeks for the set of score and loading vectors that will minimise **E**.

The second principal component, $\mathbf{C}_2$ is extracted from $\mathbf{E}_1$ and the residual matrix $\mathbf{E}_2$ is calculated from $\mathbf{E}_1$ and $\mathbf{C}_2$:

$$\mathbf{E}_2 = \mathbf{E}_1 - \mathbf{C}_2 \qquad (9)$$

The procedure can be repeated until the number of principal components equals the least of the numbers of variables or objects.

When the correlations between the variables are large, the first principal components will explain a large portion of the total variance in **M**. After extraction of principal components, the score vectors contain the information about the relationships between the objects and the loading vectors contain information about relationships between the original variables.

The original objects and variables can be investigated by plots of score vectors against score vectors and loading vectors against loading vectors. Such plots are referred to as score plots and loading plots respectively. An example is illustrated in Figure 10, where score vector 1 is plotted against score vector 2, and loading vector 1 are plotted against loading vector 2 in ordinary xy-scatterplots.

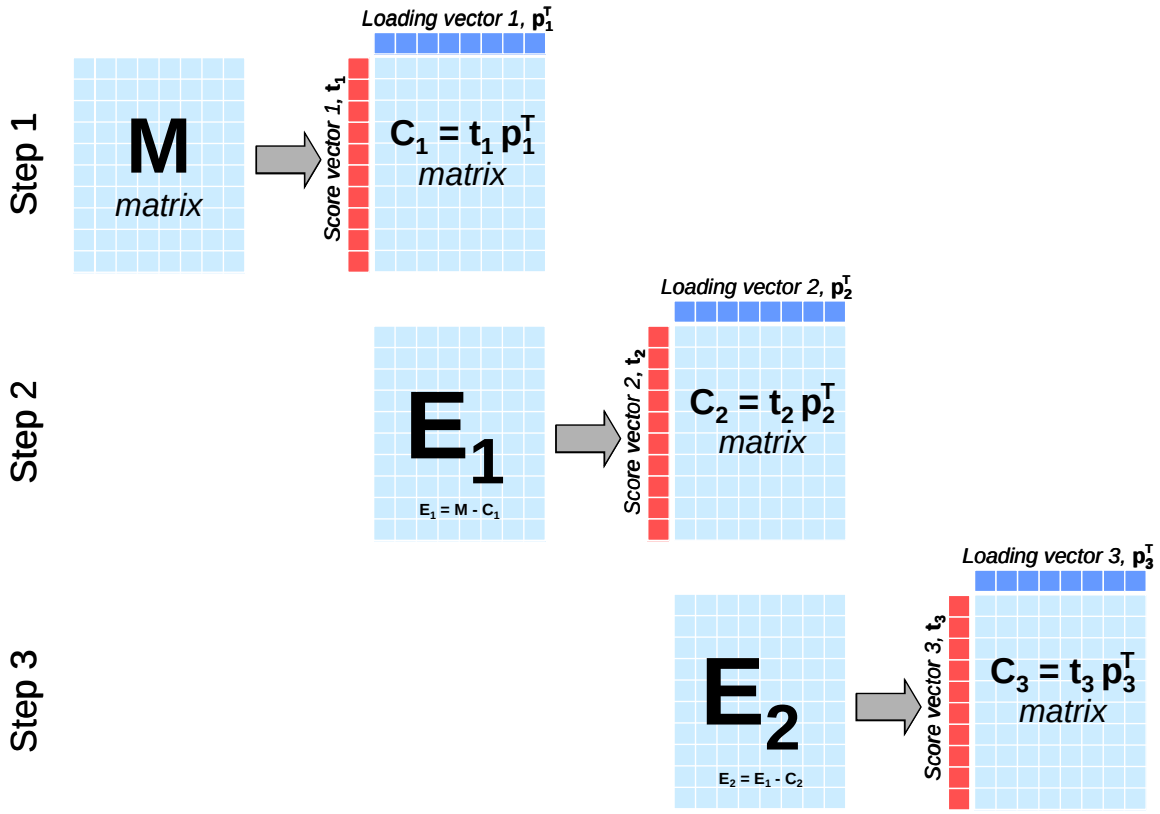



**Figure 9** *Matrix decomposition by PCA*

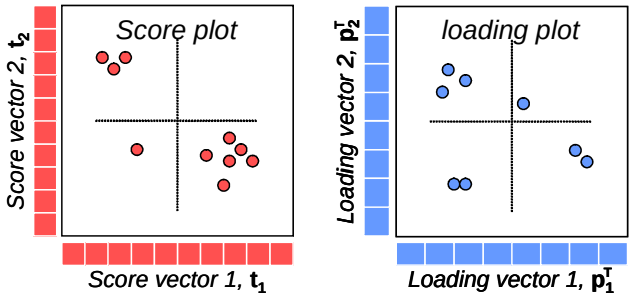



**Figure 10** *Score plot and loading plot.*

Some experience is needed to interpret PC-plots, but a few simple rules can often be applied. Assuming that the displayed principal components represent a large portion of the variance, objects that are close in score plots are also close in the original multi-dimensional space. Thus, the distance between the objects in the score plots is a measure of the similarity between objects. In the score plot in Figure 10 there are two classes of objects, five objects in the second quadrant that are similar, and a second class of objects in the fourth quadrant that are different from these. The single sample in quadrant three is not related to any of the groups.

In the loading plots, variables that lie in the same direction from the origin tend to be positively correlated, variables that lie in opposite directions are negatively

correlated, and variables that are located 90° to each other are uncorrelated. In the loading plot there are three groups of variables that are closely related. The variables in the second quadrant are negatively correlated to the variables in the fourth quadrant, while the two variables in the third quadrant are uncorrelated to these.

Usually the two first principal components are interpreted this way. When interpreting score and loading plots, it is important to consider the portion of the variance that is explained by the two principal components because important information may also be present in other components. Various validation methods indicate whether the principal components represent data structure or random noise.

### *1.4.4. Multivariate regression techniques*

Multivariate regression techniques are applied when a response variable, **y**, is to be explained from a number of x-variables ($\mathbf{x}_1$, $\mathbf{x}_{2,}$ …, $\mathbf{x}_n$), often denoted as *independent variables* or *predictors*. From algebra it is known that when the number of samples equals the number of variables, exact solutions may be found. In matrix notations this can be written as:

$$\mathbf{y} = \mathbf{Xb} \qquad (10)$$

where **y** is a vector holding the values of the response variable and **b** is the vector holding the coefficients that describe the relationship between the response vector **y** and the matrix **X**. **b** can be found by inverting the **X** matrix and multiplying by **y**:

$$\mathbf{b} = \mathbf{X}^{-1}\mathbf{y} \qquad (11)$$

When the number of samples is lower than the number of variables, the set of equations has an infinite number of solutions; the matrix **X** is not quadratic and not invertible. When there are more samples than variables the best estimate for **b** can be found by multivariate regression techniques. In ordinary multiple linear regression (MLR), **b** is found by the following equation:

$$\mathbf{b} = (\mathbf{X}^{T}\mathbf{X})^{-1}\,\mathbf{X}^{T}\mathbf{y} \qquad (12)$$

MLR has certain limitations, particularly when the number of variables is large, or when the degree of correlation between the variables or between the samples is large. When spectra are used as x-variables, the number of variables is usually large (often hundreds). In these cases the number of samples is typically lower than the number of variables and MLR gives no solution. Covariance between the variables or objects may lead to poor estimates of **b** because the matrix ($\mathbf{X}^{T}\mathbf{X}$) is rank deficient or ill-conditioned [77] (see Section 1.5.2).

The above-mentioned drawbacks of MLR may be solved by applying latent-variable techniques. In principal component regression (PCR), the **X**-matrix is first decomposed by PCA, which reduces the original number of variables to a lower number of principal components. An MLR is then based on the PCA score vectors.

An important attribute of PCR is illustrated in Figure 9. After the first principal component is extracted, PC1 is subtracted from the data matrix before PC2 is extracted. Thus, PC2 cannot contain the same information as PC1; PC3 cannot contain information explained by PC2 and so on. The covariance between principal component scores is zero (*orthogonality*); correlation between the variables is therefore no problem when the PC scores are used as variables in MLR.

PLSR is a similar technique to PCR, but there is an important difference in the extraction of the latent variables: in principal component regression, the extraction algorithm extracts latent variables that explain as much as possible of the variation in the **X**-matrix. In PLSR the algorithm extracts latent variables that explain as much as possible of the common variance between the **X**-matrix and the **y**-vector. PLSR is therefore a more powerful regression technique; fewer PLS components than PCs are often needed to obtain good estimates for **b**. PLS scores and loadings are often more easy to interpret than scores and loadings from PCA. For these reasons, PLSR are often preferred over PCR, particularly when the number of variables is large. Comprehensive reviews of PLSR are given in [77–79].

### *1.4.5. Variable weighting*

The pretreatment of variables is important in latent-variable methods. When latent variables are extracted, the variables with largest absolute variance will have the largest influence on the models. When the difference between the variables is large, a few variables will dominate the models, while others are poorly explained. These problems can be solved by proper weighting of the variables.

The most common solution is standardisation. Each variable is divided by its own standard deviation; the result is variables with standard deviations equal to one. Standardisation should be used with care. Variables with low variance often have a low signal-to-noise ratio in analytical chemistry, *e.g.* parts of spectra where there is little signal. Because of the low absolute variance, standardisation leads to multiplication by a large number and amplification of the noise.

Other weighting procedures include division by the arithmetic mean of the variable or using the logarithm or roots of each variable. Subtraction of the mean (*centring*) of each variable is normally performed before PCA or multivariate regressions.

### *1.4.6. Error estimates and validation*

The accuracy of multivariate regression models for prediction of a response variable, *y*, should be validated by a suitable method. Common validation methods are use of test sets and cross-validation techniques.

In the test set method, the dataset is separated into a calibration set (training set) and a test set. The regression model is built on the objects in the calibration set and applied for prediction of the response variable for the objects in the test set. The model is evaluated by comparing predicted ($y_p$) and 'true' ($y_t$) values of the response variable for the objects in the test set. The objects in the test set are typically randomly selected, but may also be data that are acquired at different time or at different conditions than the calibration set, *e.g.* when the stability of models are evaluated. The drawback with test set validation is that the model is calibrated and validated on only a part of the available data and the results may be highly dependent on single objects in small datasets.

In cross-validation, the objects in the original dataset are partitioned into *k* subsets. A single subset is selected as validation set and the model is calibrated on the remaining *k*–1 subsets. The model is then applied for calculation of $y_p$ for the objects in the validation subset. The procedure is repeated until all subsets have been used as validation sets and the final model are evaluated from the residuals ($y_p$–$y_t$) of all objects combined. Thus, with cross validation, all objects are utilised in validation of the model. It is often common to set *k* equal to the number of samples; this is frequently referred to as leave-one-out cross validation or full cross validation. Cross validation techniques may over-estimate the accuracy of the models in cases where the objects have a 'class like' distribution or if there are replicate objects in the dataset, because the objects in the validation set may be represented in the calibration set by their replicates or objects in the same class.

Various parameters are applied for evaluation of model performance. Bias, δ, is the average residuals:

$$\delta = \frac{1}{N}\sum_{n=1}^{N}(y_{p,n} - y_{t,n}) \qquad (13)$$

Significant bias means that the model is systematically over-estimating or under-estimating $y_p$. This may for instance occur if the model is evaluated with a test-set in cases where there has been a drift in the system.

Standard error of prediction (SEP) is the standard deviation of the residuals:

$$\mathrm{SEP} = \sqrt{\frac{1}{N\text{-}1}\sum_{n=1}^{N}(y_{\mathrm{p},n} - y_{\mathrm{t},n} - \delta)^2} \qquad (14)$$

Since there is a correction for bias in this formula, SEP will be unaffected by the bias and is a good estimate of the overall error only in cases where the bias is negligible. A more general error estimate that also accounts for the bias is root mean square error of prediction (RMSEP):

$$\mathrm{RMSEP} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(y_{\mathrm{p},n} - y_{\mathrm{t},n})^2} \qquad (15)$$

Standard error of calibration (SEC) and root mean squared error of calibration (RMSEC) are calculated by the same formulas as SEP and RMSEP, but on calibration residuals (the calibration set and validation set are identical). The correlation coefficient ($r$) or the coefficient of determination ($r^2$) for a linear regression between $y_\mathrm{p}$ and $y_\mathrm{t}$ may also serve as a rough indication of the precision of a regression model. However, this is only an indicator of correlation between $y_\mathrm{p}$ and $y_\mathrm{t}$ and will not account for bias or slopes different from 1.

## 1.5. Multivariate deconvolution methods

Even with modern capillary chromatographic columns, complete resolution of all analytes in a chromatogram may not be achieved. By using detectors that provide spectral information, *e.g.* GC–MS or liquid chromatography (LC) with diode-array UV detection, there are several possibilities for quantification of overlapping peaks. In cases of full selectivity, *i.e.* when all analytes have a signal at one or more wavelengths or ions that is absent in the other analytes in the peak cluster, quantification of the overlapping peaks is a trivial task if the selective signals have a sufficient signal-to-noise ratio.

But the information in the spectra may be used for quantification also in situations without selectivity, even with severely overlapping peaks. If standards or pure spectra of the analytes are available the problems may be solved by regression. Another solution, which does not require that reference spectra are available, is multivariate deconvolution of the overlapping chromatographic peaks.

### *1.5.1. Theory of multivariate deconvolution*

The purpose of multivariate curve resolution methods is to decompose the raw data matrix, **X**, into matrices containing pure spectra, $\mathbf{S}^\mathbf{T}$, in row vectors and pure chromatographic profiles, **C**, in column vectors (Equation 16 and Figure 11).

$$\mathbf{X} = \mathbf{C}\mathbf{S}^{\mathrm{T}} + \mathbf{E} \quad (16)$$

**X** has the dimension $N \times M$, **C** has the dimension $N \times A$ and $\mathbf{S}^{\mathbf{T}}$ has the dimension $A \times M$. $N$ is the number of data points in the chromatographic profile (number of spectra measured) and $M$ is the dimension of the spectra (number of wavelengths or masses). $A$ is the number of analytes in the peak cluster. **E** is the error matrix, which has the same dimension as **X**. Each analyte in the system is represented by a column vector, **c**, describing the chromatographic profile and a row vector, $\mathbf{s}^{\mathbf{T}}$, describing the spectrum, and Equation 16 can be rewritten as:

$$\mathbf{X} = \sum_{a=1}^{A} \mathbf{c}_a \mathbf{s}_a^{\mathbf{T}} + \mathbf{E} \quad (17)$$

Assuming that **E** is small enough to be neglected, estimates of **C** can be found from estimates of **S** by the following equation:

$$\mathbf{C} = \mathbf{X}\mathbf{S}(\mathbf{S}^{\mathrm{T}}\mathbf{S})^{-1} \quad (18)$$

Similarly, estimates of $\mathbf{S}^{\mathbf{T}}$ can be found from estimates of **C**:

$$\mathbf{S}^{\mathrm{T}} = (\mathbf{C}^{\mathrm{T}}\mathbf{C})^{-1}\mathbf{C}^{\mathrm{T}}\mathbf{X} \quad (19)$$

The estimates of **C** and $\mathbf{S}^{\mathrm{T}}$ can be obtained by several approaches, which roughly can be divided into direct and iterative methods. In the direct methods the pattern of peak overlap is typically analysed by evolving factor analysis (EFA) [80,81], modifications of EFA, such as the fixed-size moving window EFA [82] or eigenstructure tracking analysis [83]. Other methods, such as latent projective graphs (LPG) [84], may also be applied. The overlap pattern reveals where pure spectra or chromatographic profiles can be found, and complete resolutions can often be achieved by the application of Equation 18 or 19 or by using information from zero-concentration windows to find estimates for **C** and $\mathbf{S}^{\mathbf{T}}$. This can be done by rotating the scores and loadings from PCA into estimates of **C** and $\mathbf{S}^{\mathbf{T}}$:

$$\mathbf{C}\mathbf{S}^{\mathrm{T}} = \mathbf{T}\mathbf{R}\mathbf{R}^{-1}\mathbf{P}^{\mathrm{T}} \quad (20)$$

**R** is a $A \times A$ rotation matrix, **T** is the PCA score vectors and has the same dimension as **C**. $\mathbf{P}^{\mathrm{T}}$ is the PCA loading vectors with the same dimensions as $\mathbf{S}^{\mathrm{T}}$. Various procedures for finding **R** are described and compared elsewhere [80,81,84–86].

In the iterative methods initial estimates of **C** or $\mathbf{S}^{\mathrm{T}}$ are refined in a repetitive manner until a convergence criterion is met. The equations above (18–20) are often involved in one or several steps in the procedure. Good initial estimates, which can be obtained by several methods [87–92], are essential. The initial estimates are refined by applying constraints on the estimates of **C** and $\mathbf{S}^{\mathrm{T}}$. Common constraints are positivity in **C** and $\mathbf{S}^{\mathrm{T}}$ and unimodality in **C**, *i.e.* each chromatographic peak in **C** has only one maximum. Common iterative procedures are alternating regression

[93], iterative target transformation factor analysis [94] and Gentle [95]. If available, information from several samples can be combined to give more accurate results [96,97].

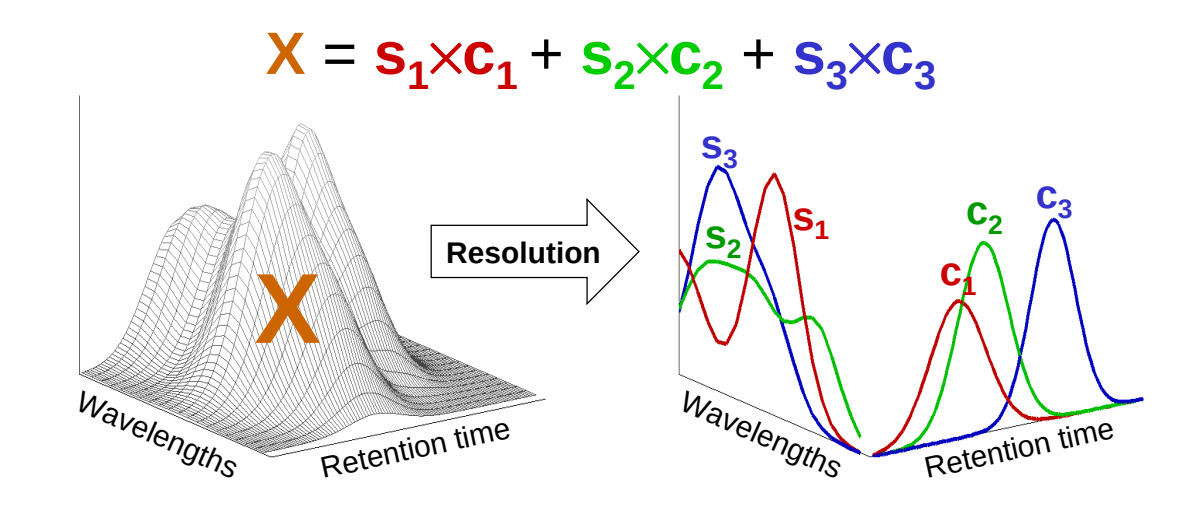



**Figure 11** *The purpose of multivariate deconvolution is to find the pure chromatographic profiles, **C**, and pure spectra, **S**, from raw-data, **X**.*

### *1.5.2. Determination of the number of components*

The mathematical definition of the *rank* of a matrix is the number of linearly independent rows or columns. Taking the noise into consideration, a more practical definition is that the rank equals the number of factors that are significantly different from the noise. The rank of the raw data equals the maximum number of spectra and chromatographic profiles that can be resolved from a peak cluster by a self-modelling deconvolution method. Ideally, the rank of the raw data matrix, **X**, should therefore be equal to the number of analytes. Several factors may reduce the rank of the system and hinder successful resolution. The rank is reduced if there are compounds with identical spectra or spectra that are linear combinations of other spectra. The rank is also reduced if two chromatographic peaks are completely overlapping. Consequently, there must be a certain chromatographic resolution, and a certain difference between the spectra to achieve successful deconvolution.

A large number of methods are used to estimate the number of significant factors of a data matrix. The various methods are thoroughly reviewed elsewhere [75]. There is probably no general method that can be applied for estimation of the number of compounds in GC–MS peak clusters because the noise patterns differ between instruments and between different applications on the same instrument.

### *1.5.3. Noise*

In any real chromatographic data, various types of noise will be present and may hinder successful resolution. Noise in chromatographic systems is typically heteroscedastic; the noise increases with signal strength. The expected accuracy and precision of a chromatographic peak are commonly expressed in per cent of the peak area, meaning that the noise is expected to increase proportionally with the peak size. The heteroscedastic noise may give inaccurate results when peaks with large differences in size are resolved. Because noise from the large peaks in a system will also influence the smaller peaks, the amount of noise relative to the peak size of smaller peaks will be large, and lead to inaccurate quantification of these. Thus, the limits of detection or accurate quantification of an analyte will not only depend on the peak size of the analyte, but will also depend on the size of other peaks in the cluster.

Another effect that may limit the possibility of resolution is the *scan-effect* [98,99, Paper IX]. As can be seen from Equation 17, the principles of multivariate deconvolution methods require that the spectrum of an analyte is described by a single vector, $\mathbf{s}^T$, and the spectrum should therefore be constant during the elution of a pure peak. Quadrupole and sector mass spectrometers are scanning instruments, where the ions are recorded in sequence. The concentration in the detector changes during the time interval between the first and the last ions recorded in a scan. This leads to ‘false’ correlations between ions that are close in the scanning sequence. The scan effect may lead to overestimation of the number of components in a peak cluster (the rank) and may also hinder successful resolution of peaks with similar spectra, even if the number of compounds is correctly estimated. Possible solutions for reduction of the scan effect are given in Paper IX.

# 2. Identification of FAME from shifts in ECL-values

This chapter is a discussion of Papers I and II, which both deal with the use of temperature induced shifts in ECL values for the identification of FAME. The strong dependence of the polarity of the cyanopropyl phases on temperature is an essential issue in both papers.

## 2.1. Temperature induced shifts in ECL-values

A method for the identification of fatty acid methyl esters based on the analysis of shifts in ECL values is described in Paper I. It has been shown that the polarity of cyanopropyl phases is temperature dependent [17] and that ECL-values of unsaturated fatty acids increase with temperature on these columns [100–102]. In temperature-programmed GC, increased temperature gradients or decreased carrier gas flow will increase the elution temperature of a compound, and therefore have similar effects as increased temperature in isothermal chromatography. This can be seen in the chromatograms in Figure 12, which compares the elution patterns of FAME analysed with two different programs on the BPX-70 phase. The difference in ECL value for the highly unsaturated fatty acid 22:6 *n*–3 is as high as 0.45 ECL units. The differences for less unsaturated fatty acids are smaller, only 0.10 ECL units for 24:1 *n*–9.

In the work described in Paper I, the effects of varying start temperature, temperature gradients and column flow in linear temperature programs on the BPX-70 phase were analysed. The factors were varied in a full factorial $3^3$-design as shown in Figure 13. Chromatographic parameters are described in detail in Paper I. PCA (with programs as objects and fatty acids as variables) of the ECL values of 27 FAMEs with different chain lengths and number of double bonds showed that the two first principal components explained 99% of the variance and that four programs adequately span the variation in the dataset (See Figure 1 in Paper I). The four programs and the centre point in the design were therefore selected as a suitable set of programs and applied for further studies in Paper I–IV. These programs are referred to as the *standard programs* in the remaining discussion.

PCA on ECLs acquired with the standard programs, with the fatty acids as objects and the programs as variables, gave score plots where the fatty acids were distributed according to the chain lengths and number of double bonds. An example from Paper I is shown in Figure 14. The pattern in this figure is similar to the retention maps of FAME that are acquired by two-dimensional gas chromatography (GC × GC) [103]. In Paper I it is also shown that the number of carbons and the

number of double bonds in FAME can be determined by multivariate regression on the ECL data from the five standard programs.

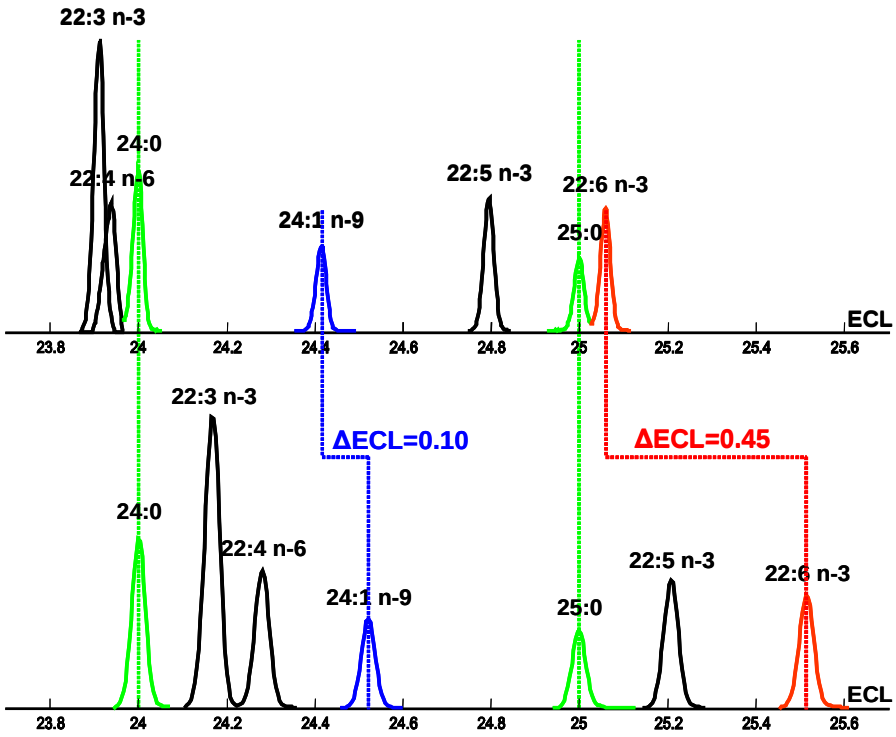



**Figure 12** *Elution pattern from ECL 23.8 to 25.6 with two different temperature and flow programs on BPX-70. Upper chromatogram: Start temp. 160 °C, temp. gradient 2 °C/min, column flow 26 cm/sec; lower chromatogram: Start temp. 190 °C, temp. gradient 4 °C/min, column flow 18 cm/sec. Further details are given in Paper II.*

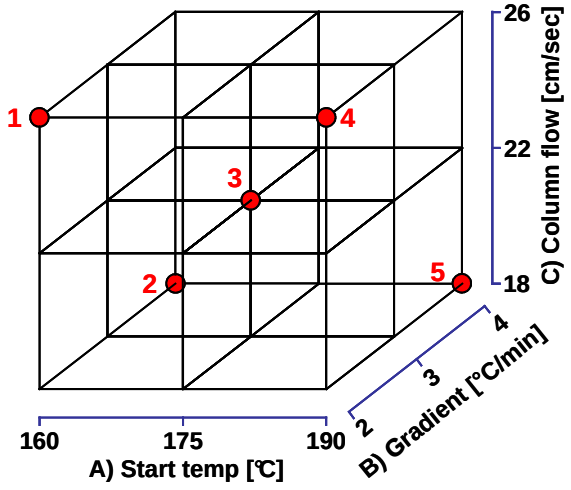



**Figure 13** *Levels used in $3^3$ factorial design used in the study of ECL-shifts on a 60 m BPX-70 column. Red spots mark the standard programs (1–5) that were selected for further studies in Paper I–VI.*

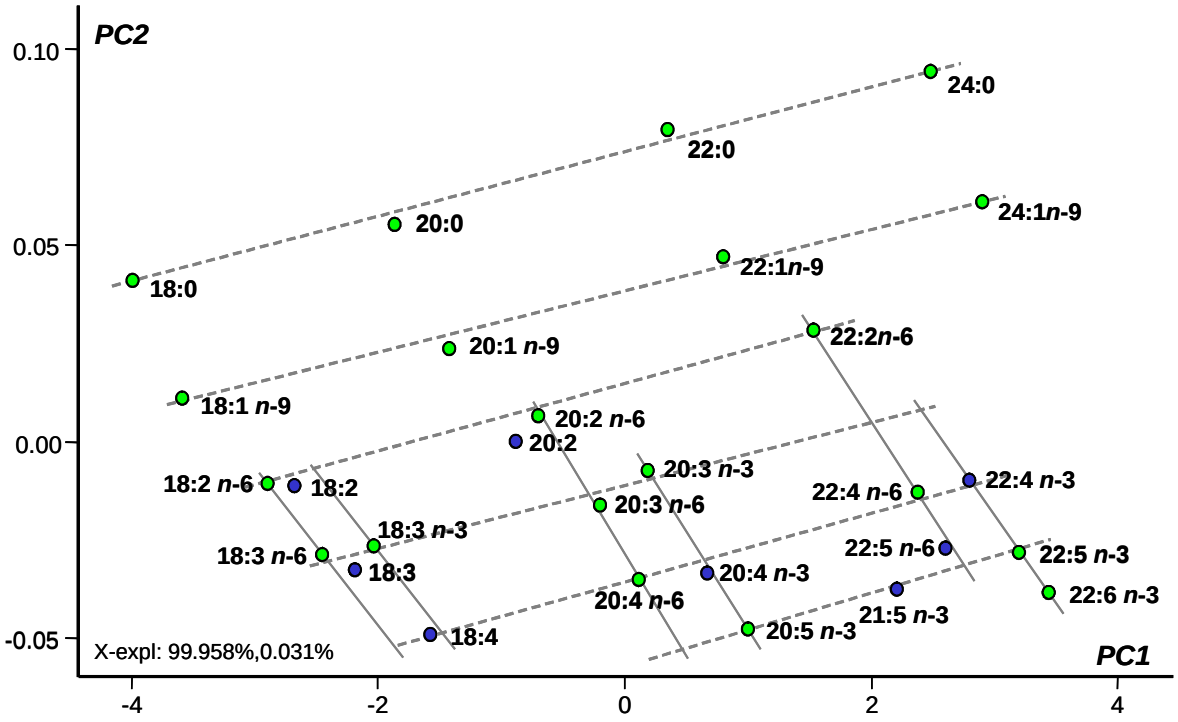



**Figure 14** *PCA score plot of ECL values of C18–C24 FAMEs as objects. Variables were the five standard programs. Broken lines indicate the number of double bonds; solid lines indicate the n–3 and n–6 series. Green circles are fatty acids in the reference mixture GLC-461. Blue circles are tentatively identified marine fatty acids. PCA was performed on unweighted and mean centred variables. Further details are given in Paper I.*

## 2.2. Two-dimensional fatty acid retention indices

A problem with identification based on plots like Figure 14 is the lack of stability. Principal components are ‘abstract factors’ that explain the properties of a given object relative to the other objects in the dataset, and with scales that has no direct chemical meaning. Thus, the score values of each fatty acid on the two components (PC1 and PC2) depends on which fatty acids that are present in the dataset, and will also vary with drift in stationary phase properties and other chromatographic conditions. Because of this instability it is difficult to identify compounds by comparisons with previously acquired retention data.

The solution is to stabilise the system by defining retention indices in two dimensions; two-dimensional fatty acid retention indices (2D-FARI) were introduced in Paper II. Traditional retention index systems like Kovats indices or ECL values are typically based on a series of homologues where the only difference between the references is a single property, normally the chain length of the compounds. In a two-dimensional retention index system, the references must also vary in a second property. For fatty acids, an obvious choice would be to base the

system on the number of *cis* double bonds in addition to the chain length. These properties are also the two effects that are best explained by the plot in Figure 14.

The reference mixture GLC-461 spiked with 22:3 *n*–3 and additional saturated FAMEs was used as calibration sample and analysed with the same standard programs that were selected in Paper I (Figure 13). ECL values for the standard programs were used in multivariate regression models for prediction of the chain length and the number of double bonds in the FAMEs in the calibration sample. The *predicted* values from these multivariate models were then applied to define the two retention index scales, $FARI_A$ and $FARI_B$. The values that define the retention indices are given in Table 2 in Paper II. Because the values that define the scales are predicted values, they are not exactly the chain length and the number of double bonds, but the values explain these two factors if rounded to the nearest integers. Predicted values from the multivariate regression model were applied to define the scales, instead of true values, because values that can be accurately explained by the ECL data from the standard programs will give a more stable system.

2D-FARI values are acquired by analysing the reference sample and other samples using the five standard programs. Multivariate regression is then applied to find the relationship between the ECL data for the standard programs and the defined 2D-FARI values for the compounds in the reference mixture. The regression models are thereafter applied to predict 2D-FARI values for other FAMEs from their ECL values acquired using the five programs. This procedure aligns all FAMEs relative to the reference compounds and the results can be presented by ‘maps’ as shown in Figure 15.

In the 2D-FARI map in Figure 15 the fatty acids are positioned within squares according to the chain length and the number of double bonds. However, there are exceptions for FAMEs with *trans* double bonds and FAMEs with a terminal double bond (*n*–1). The special behaviour of *trans* fatty acids on CNP phases is discussed further in Section 3.

The polarity of CNP phases will often decrease with time. The length of a capillary column may also decrease with time because it may be necessary to cut a piece of the column each time the column is installed in the GC because the first meters of the columns may degrade if ‘dirty’ samples are injected. Thus, old columns may be several meters shorter than new columns, and will therefore have higher carrier gas velocity than new columns when used with identical injector pressures.

The data shown in Figure 15 are from two BPX-70 columns with different properties; the difference in ECL for 22:6 *n*–3 acquired on the two columns is 0.08 to 0.09. This is approximately equal to the width of a peak at the baseline. It is shown in Paper II and in Figure 15 that the 2D-FARIs are relatively robust towards differences in column properties and also that the 2D-FARI data can be estimated from programs with different temperature and flow settings. The 2D-FARI data for

the BPX-70 phase could also be estimated from ECL data acquired on a more polar CNP phase (SP-2560) but with some reduction in accuracy.

With ECL values and other one-dimensional retention indices, changes in chromatographic conditions will have largest effect on the most unsaturated compounds (see Figure 12). 2D-FARIs are robust towards these changes because information about the polarity (number of double bonds) of the compounds is present in the data matrix applied in the regression. This information will not be present in ECL data from a single chromatographic run and a similar robustness towards changes in the chromatographic conditions cannot be achieved from ordinary univariate retention data.

It may seem unnecessary and inconvenient to use as many as five different GC programs to calculate the two-dimensional data. It was shown in Paper I that the information in the dataset from the five programs can be explained (99.99%) by two principal components and that the information about the chain length and number of double bonds in FAME can be achieved by comparing the ECL data from only two programs. Thus, it can be argued that there are only two significant dimensions in the data. However, using more programs will increase the stability of the method. In complex mixtures there is also a risk of chromatographic overlap that may hinder accurate calculation of the ECL data. If a compound overlaps in one or more chromatograms it will still be possible to calculate accurate 2D-FARI data based on the remaining programs. More than two programs also allow outlier detection by multivariate methods like PCA.

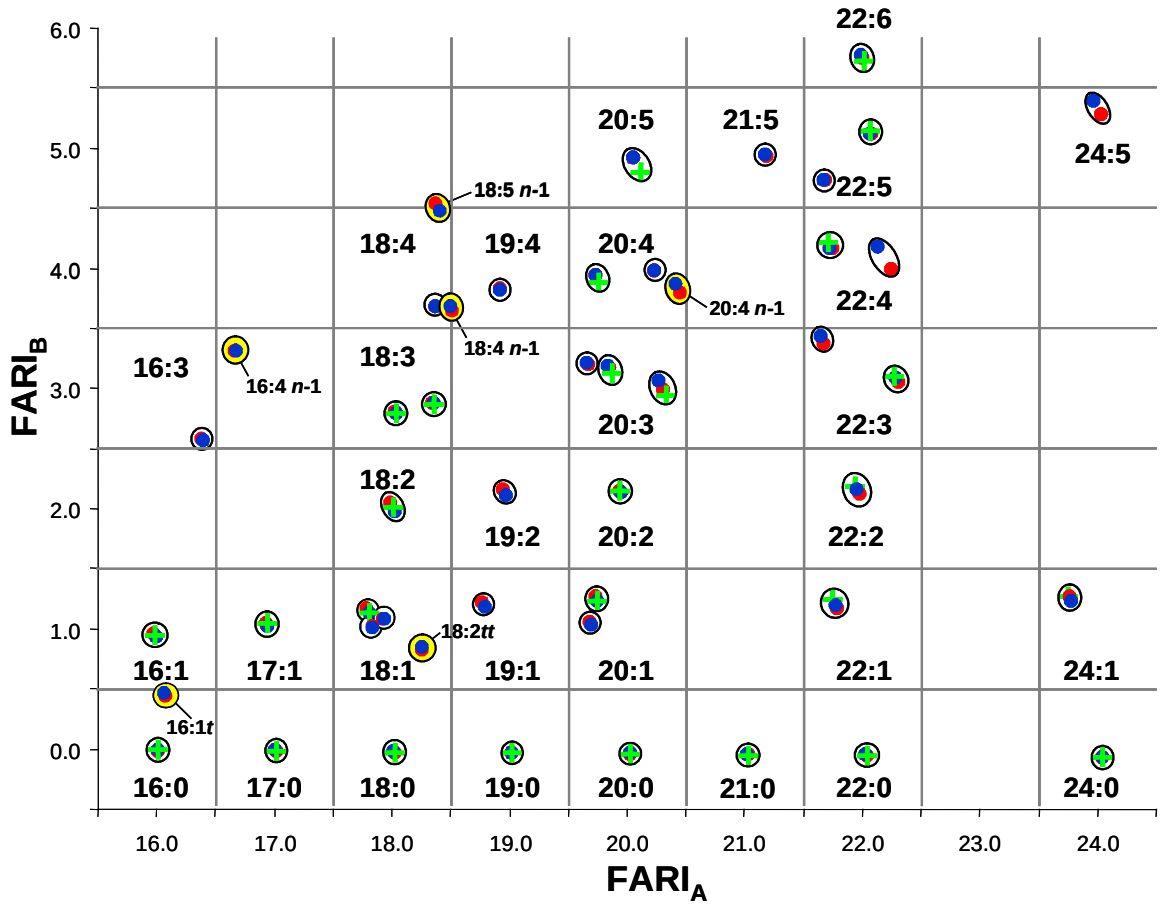



**Figure 15** *2D-FARI of FAME analysed on two different BPX-70 columns. The values are the same as shown in Paper II (Figure 5). Green crosses mark values that are given by definition for FAMEs in the calibration sample. Red and blue spots mark results acquired on 'new' and 'old' column, respectively. Trans fatty acids and n–1 fatty acids are shown on yellow background.*

# 3. Application to *trans* isomers

This section is basically a discussion of Papers III and IV, but also includes some data from Paper I and II. The geometries of the double bonds in PUFA are essential for the chromatographic behaviour of the compounds. It is often assumed that *trans* isomers elute before the corresponding *cis* isomers on highly polar CNP phases. This assumption seems valid for monoenes, dienes and most tetraenes. However, the picture is more complicated with more unsaturated PUFAs.

The effect of increased temperature on CNP phases is much less for *trans* double bonds than for *cis* double bonds. This has been shown for monoenes [101] and can also be seen by comparing ECL values from Paper III for *trans* and *cis* isomers of EPA and DHA with *cis* monoenes for the two standard programs referred to as Program 1 and 5 (See detailed discussion in Paper III).

The different behaviour of *cis* and *trans* double bonds has implications both for the identification of *trans* fatty acids from retention data and for optimisation of chromatographic elution patterns of *trans* fatty acids. A PCA score plot of ECL data of *cis* and *trans* isomers of 18:1 *n*–9, 18:2 *n*–6, 18:3 *n*–6 and 18:3 *n*–3 that were acquired with the standard programs is shown in Figure 3 in Paper I. In this case, there is no variation in the chain length in the dataset and the fatty acids are distributed according to the number of *cis* and *trans* double bonds.

The effect of *trans* double bonds is also visible in the 2D-FARI map in Figure 15 where the $FARI_B$ values of *trans* 16:1 and all-*trans* 18:2 indicate that the effect of a *trans* double bond is slightly less than half the effect of a *cis* double bond. 2D-FARI data for all-*cis* and all-*trans* EPA and DHA are given in Paper III; the $FARI_B$ values of the all-*trans* isomers were only 27 and 28 % of the values for the all-*cis* isomers.

## 3.1. Retention data for *trans* isomers of EPA and DHA

In Paper III the different behaviour of *cis* and *trans* double bonds were utilised for optimisation of the elution patterns of isomers of EPA and DHA with one and two *trans* double bonds (1-*trans* and 2-*trans* isomers). These highly unsaturated fatty acids are abundant in fish oils and other marine lipids. Geometrical isomerisation of double bonds occurs at high temperature processing of edible oils and the isomerisation rates increase with the number of double bonds [104,105]. Geometrical isomers of these fatty acids have been reported in processed fish oil [106,107] and also in rat liver as a consequence of metabolism of *trans* isomers of 18:3 [108,109].

Accurate quantification of *trans* isomers of EPA and DHA requires that the *trans* isomers are chromatographically separated from the all-*cis* isomers. The geometries of double bonds have limited influence on the retention in GC, and with the high number of possible isomers it is likely that one or several *trans* isomers will overlap with the all-*cis* isomer. The *trans* isomers can be separated from the all-*cis* isomers by LC or thin-layer chromatography (TLC) on stationary phases impregnated with silver ions. However, the main purpose of the work reported in Paper III was to search for resolution windows on CNP phases that allow direct analysis of the *trans* isomers by a single GC analysis.

The all-*cis* isomer and isomers with one and two *trans* double bonds of EPA and DHA were analysed with the five standard programs. The relationships between the ECL values of the all-*cis* isomers and the *trans* isomers were found by linear regression and the regression lines are shown in Figure 16. As can be expected, the ECLs of the *trans* isomers increase with the ECLs of the all-*cis* isomers, but with less steep slopes. It can also be seen that the slopes of isomers with the same number of *trans* double bonds are nearly parallel.

The slopes in Figure 16 show that it is possible to move the all-*cis* isomers relative to the *trans* isomers by changing the temperature and flow conditions, but isomers with the same number of *trans* double bonds cannot be moved relative to each other. The largest resolution is found where the distance between the red line (representing the all-*cis* isomer) and the other lines is maximised.

Sufficient resolution ($R_s = 1$) requires a distance between the isomers of approximately 0.05 ECL units (see Paper III for details). There exist no windows that will provide this resolution between the all-*cis* and the 2-*trans* isomers. However the amounts of 2-*trans* isomers are formed in a much lower rate than 1-*trans* isomers and the amounts in real samples can be expected to be negligible [107,110].

The best resolution between all-*cis* EPA and the 1-*trans* isomers is in the area between Program 1 and 4. Program 1 also provided the best resolution of all-*cis* and 1-*trans* DHA, but the regression lines indicate that resolution will be better for programs with lower ECL value for the all-*cis* isomer. This can be achieved by lowering the temperature gradient or increasing the column flow. However, it should be emphasised that the selection of chromatographic conditions is often a trade-off between separation efficiency and the time required for the analysis. Lower temperature gradients will increase the analysis time; too high carrier gas flow will give loss of column efficiency and reduced sensitivity for MS detectors.

The properties of the EPA and DHA isomers were also investigated on SP-2560, which showed similar correlations between the all-*cis* and the *trans* isomers (see Paper III for details). The most notable differences were that SP-2560 separated four 1-*trans* EPA isomers and that it was not possible to achieve sufficient resolution

between the all-*cis* isomer and the first 1-*trans* isomer eluting after the all-*cis* isomer. It was therefore concluded that BPX-70 is a more suitable column for analysis of these isomers. However, in subsequent studies on thermally isomerised EPA and DHA concentrates [110], it was found that only minor amounts of this isomer is formed and that there is also a problematic overlap between 20:4 *n*–3 and the first 1-*trans* EPA isomer on BPX-70 (Figure 17). The overlap between 20:4 *n*–3 and the 1-*trans* EPA isomer is difficult to avoid by changing the chromatographic parameters because the two compounds have the same number of *cis* double bonds.

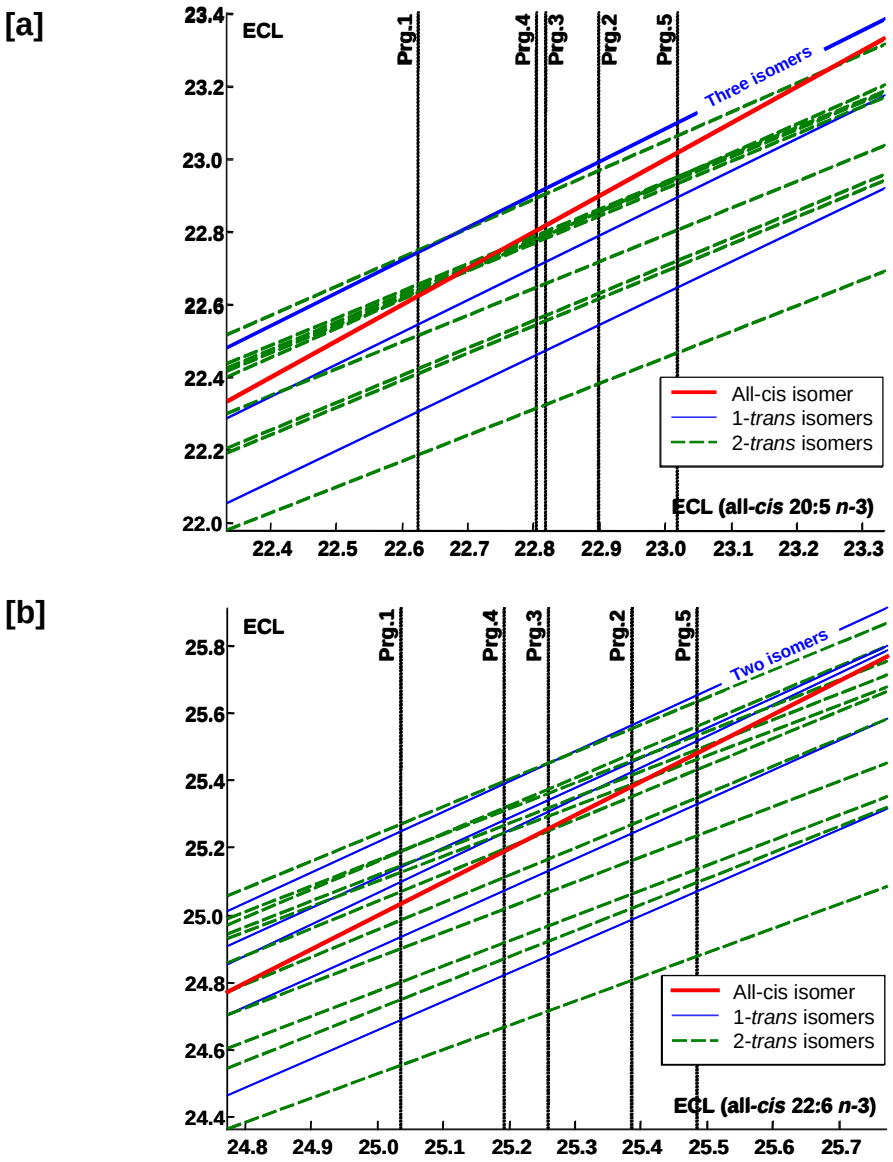



**Figure 16** *ECL of geometrical isomers of 20:5 n–3 [a] and 22:6 n–3 [b] (vertical axis) as functions of the ECL of the all-cis isomers (horizontal axis) on a BPX-70 column. Vertical lines mark the ECL values of the all-cis isomers acquired with the standard programs. Background data and equations for the regression lines are given in Paper III.*

## 3.2. Application of the retention data for *trans* isomers of EPA and DHA

The data for the retention behaviour of the *trans* isomers of EPA and DHA found in Paper III can be applied both for identification of the isomers and for optimisation of chromatographic conditions. Paper IV is about identification of *trans* isomers formed as artefacts in acid catalysed preparation of FAME from marine tissues. Both the regression lines and the 2D-FARI data from Paper III were applied for identification of the artefacts.

Because Figure 16b suggests that the best resolution is achieved with low ECL values for all-*cis* DHA, a temperature program with a low temperature gradient (1.0 °C/min) was selected. This gave ECL for all-*cis* EPA of 22.49 and all-*cis* DHA of 24.85. Even though the applied program gave ECL values outside the region investigated in Paper III, the predicted retention times for the *trans* isomers (converted from ECL values) show fairly good match with the experimental retention of the 1-*trans* reference compounds (Figure 2 in Paper IV). The regression lines in Figure 16 can therefore be extrapolated to areas outside (but near) the investigated region with acceptable accuracy.

The data from Paper III have also been applied for identification of isomers formed in heated EPA and DHA ethyl ester concentrates [110] Chromatograms of neat and thermally isomerised EPA and DHA are shown in Figure 17 together with 1-*trans* isomers. DHA was analysed with the same conditions as applied in Paper IV and ECL of the all-*cis* DHA isomer was similar as in Paper IV (24.85). EPA was analysed with a program expected to give optimal resolution of the all-*cis* and 1-*trans* isomers with ECL of the all-*cis* EPA of 22.63. There were good match between predicted and experimental retention times for the 1-*trans* isomers and there were also correspondence between predicted retention times for 2-*trans* isomers and minor compounds formed by thermal isomerisation (Figure 17).

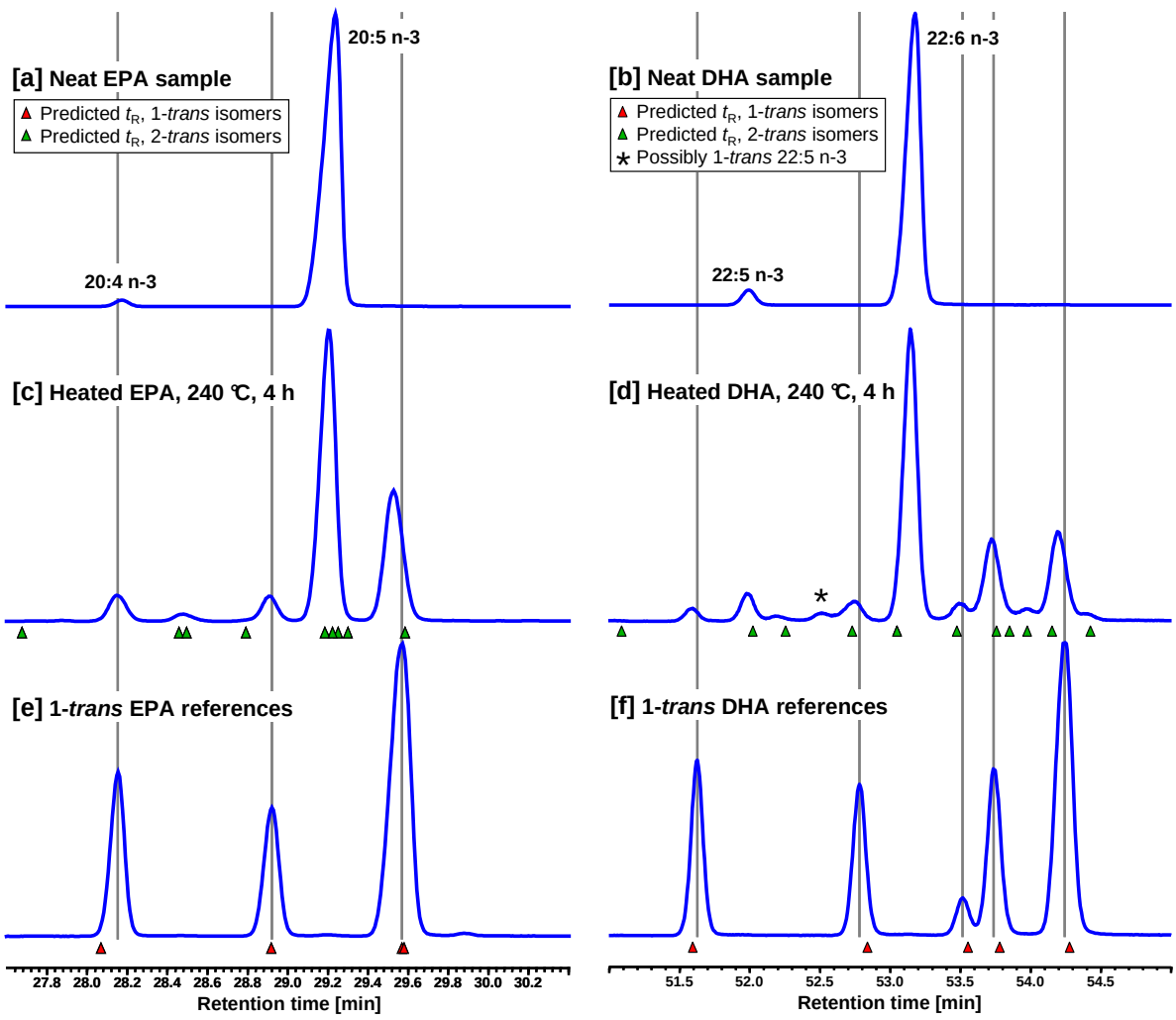



**Figure 17** *EPA and DHA concentrates analysed on BPX-70. [a] and [b] are Neat concentrates, [c] and [d] are thermally isomerised concentrates. [e] and [f] are 1-trans references. Triangles mark predicted retention times based ECL of the all-cis isomers and the equations given in Paper III. The GC temperature programs was similar to the standard programs but with start temperature of 180 °C (EPA isomers) and 160 °C (DHA isomers), gradients of 1 °C/min and carrier gas velocity of 26 cm/sec [110].*

# 4. Prediction of equivalent chain lengths

This chapter is a discussion of Papers V–VI, which are about the prediction of ECL values. There are basically two reasons why prediction of retention indices is of interest. Accurate prediction of retention indices may be a valuable tool for identification of unknown compounds that is not available as standards. Since several compounds may have identical or near identical retention times, matching retention indices will not confirm the identity of a compound. However, models that predict retention indices (or retention times) may be an effective tool for elimination of alternatives or incorrect tentative identifications.

A second reason to develop models for prediction of retention indices is related to optimisation of elution patterns. In complex samples like marine fatty acids there is large risk of chromatographic overlap and suitable reference samples that contain all compounds of interest may not be available. With models that predict retention indices it is possible to test whether a given compound will be resolved or hidden under other peaks.

Various approaches have been applied for the prediction of retention indices. These can roughly be divided into three classes:

Type I: Models based on experimentally determined solute-solvent interactions.

Type II: Models based directly on the molecular structure or parameters calculated from the molecular structure.

Type III: Models based on group additivity.

The first type of models are based on sets of *solute descriptors*, which are experimentally determined parameters describing the properties of each analyte, and experimentally determined *phase constants* (or *solvent descriptors*), which describe the properties of the stationary phases. Retention indices can be estimated by equations as shown below:

$$I = a \times A + b \times B + c \times C \ldots + n \times N \qquad (21)$$

where $a$, $b$, $c$, …, $n$ are solute descriptors and $A$, $B$, $C$, …, $N$ are phase constants. Thus, there is one solute descriptor matching each phase constant. Typical models include 3 to 6 parameters of each type [111,112]. The models may also be expanded by inclusion of other parameters describing the solute properties, like boiling points or the retention index on squalane [111,113]. The most common systems of solute-solvent interactions are the concepts of Rorschneider [113,114], McReynolds [115] and Abraham [112].

This type of models has several limitations. The main drawback is that the data required for the calculations may not be available. The solute descriptors are available mainly for relatively small molecules. The use of phase constants is also questionable; as shown in Section 2, the properties of a phase are not necessarily constant.

The alternative to models based on experimentally determined solute and solvent descriptors is to base the models directly on a description of the molecular structure or on parameters estimated from the structure (Type II). Mathematical models of the relationships between the retention properties and the structures of compounds are often referred to as quantitative structure-retention relationships (QSRR). The challenge in QSRR is often to find relevant mathematical descriptions of the structure of a molecule that can be applied as variables in the models. A mathematical description of the structure is possible for simple systems like unbranched fatty acids with methylene-interrupted double bonds. Other simple systems that can be described in a similar way are dioxins [116] polychlorinated biphenyl ethers [117] and simple hydrocarbons [118]. However, with more complicated molecules with several functional groups, or for general models that cover a wide range of functional groups and possible structures, a relevant mathematical description of the molecule is not possible. This problem may be circumvented by basing the models on suitable molecular descriptors, such as molar volumes, electron donating properties, polarizability and dipole moments that is estimated from quantum chemical calculations. A large number of parameters are typically calculated in the initial stage and several hundred parameters are often evaluated before the models are refined by a suitable variable selection procedure [119].

Models based on molecular structure will usually not apply any descriptors of the stationary phase properties, but will instead use a calibration set of compounds with known retention indices and a multivariate regression method (*e.g.* PLSR or neural networks) to find the relations between the molecular descriptors and the retention indices. The multivariate model is then applied to predict the retention indices of other compounds from their structure. Thus, the only experimentally determined parameters in these models are usually the retention indices of the calibration compounds. The drawbacks of these models are that the accuracies, especially of models that cover a large range of functional groups, are often limited, and that the calculations of the descriptors and the final models require computational power.

Retention indices may also be estimated by assuming functional group additivity (Type III), *i.e.* the presence of a functional group will have the same effect in different molecules. Models of this type are usually based on the number of carbons and contributions from ‘group retention factors’ [120–122]. Estimation of FCL values of PUFAs by summing the FCL values of monoenes with the double bonds in the corresponding positions [33–35,123] can be classified as a Type III approach.

It should be emphasised that there are approaches that do not fit nicely into the three categories described above, and there are also examples that can be described as hybrids, *e.g.* models that include both experimentally and theoretically derived molecular descriptors.

Two different approaches for prediction of ECL values have been evaluated in Paper V and VI. The work in Paper V predicts the ECL of PUFA from simple descriptors of the molecular structure and can be described as a Type II approach. The work in Paper VI predicts ECL from 2D-FARI values, which can be described as experimentally determined solute descriptors. However, the models are based on regressions on a calibration set instead of stationary phase descriptors and therefore have elements from both the type I and type II approaches.

## 4.1. Prediction of ECL values of MI-PUFA from the molecular structure

In Paper V multivariate regression models were applied to predict ECL values for methyl esters of PUFA. In PUFA with MI double bonds of *cis* geometry, the molecular structure can be described by only three parameters: the chain length, the number of double bonds and the position of the double bond system. It is well known that all three parameters will influence the ECL values of the PUFA and it was evaluated whether precise models for prediction of ECL and FCL values could be based solely on these three parameters.

As shown in Section 1.1, the positions in the double bond system in MI-PUFA can be given both as the distance from the carbonyl group (Δ-position) and as the distance from the methyl end (*n*-position). It is also well known from the studies of monoenes that the relationships between double bond positions and ECL values are not linear [32,124,125]. Shifts in positions have limited effect close to the centre of the carbon chain, while the effects increase substantially as the double bonds approaches either of the ends. A common trick to handle such non-linearities by linear methods is to include higher order terms as variables [77].

Both Δ- and *n*-positions and the higher order terms of Δ and *n* were therefore included in the models as separate variables and the performance of models based on various selections of the variables was investigated. For general models it was necessary to include both the Δ-position and the *n*-position and the higher order terms $\Delta^2$, $\Delta^3$, $\Delta^4$ and $n^2$. For models restricted to *n*–3 and *n*–6 isomers it was not necessary to include the *n*-position among the variables. The highest residuals for the most accurate models were below 0.06 ECL units, and RMSEP was below 0.030.

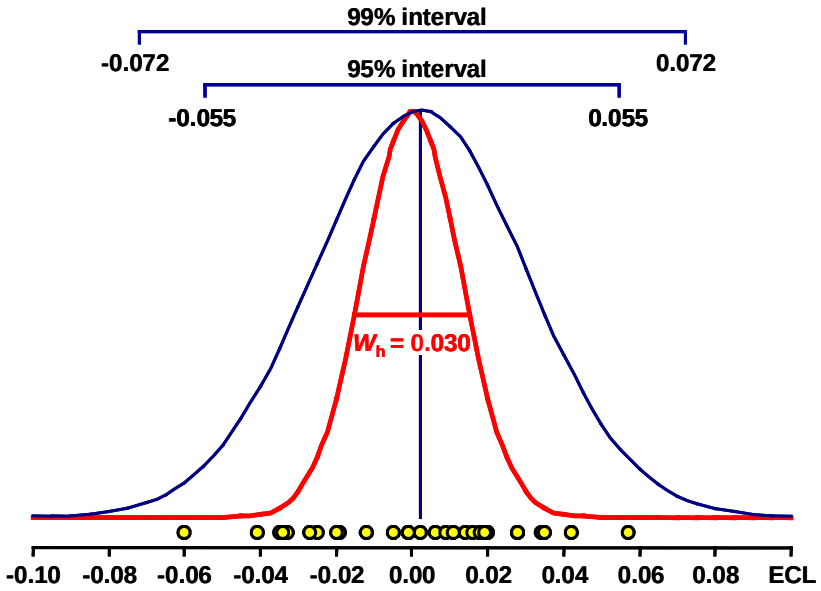



**Figure 18** *Distribution of the residuals (blue curve) compared to the chromatographic peak width (red curve) for model 19e in Paper V. The peak width is estimated from 20:5 n–3. Residuals of the single fatty acids are shown with yellow dots. The intervals indicate 95 and 99 % of the area of the error distribution curve. RMSEP = 0.027.*

There was a good correlation between predicted and experimental ECL values with slopes near one and $r^2$ of 0.9999. However, practical applications of models for prediction of retention indices require very high accuracy to be of any use and the accuracy are better illustrated if the residuals are compared to the width of a peak, as in Figure 18. Some of the residuals are outside the area covered by the estimated chromatographic peak width (red curve), which means that a predicted ECL value for a given compound may be outside the area covered by the chromatographic peak. The intervals of the error distribution curve may be used to set limits for rejection of tentative identifications. In the case shown in Figure 18, the identification of a compound may for instance be rejected if the ECL of the chromatographic peak is more than 0.072 or 0.055 units from the predicted value (99 and 95% intervals, respectively).

Both ECL and FCL values were applied as dependent variables in the models in Paper III. The residuals were almost identical. Different regression methods were also applied. PCR and PLSR, methods based on latent variables, gave approximately equal RMSEP, while MLR gave poorer predictions.

The regressions were tested on ECL data from three different GC programs (standard programs 1, 3 and 5). The residuals of models based on different GC programs were correlated, *i.e.* fatty acids with a large positive residual on models based on one program had large positive residuals on models based on the other two programs. These correlations should not be present if the accuracy of the ECL data

was the limiting factor for the accuracy of the models. Because there is systematic variation in the ECLs that is not explained by the best models, it should be room for further improvements.

## 4.2. Prediction of ECL values from 2D-FARI data

Since the two-dimensional fatty acid retention indices are calculated from the ECL values acquired at different chromatographic conditions, the two values contain information about how the ECL values will change with variations in chromatographic parameters. Prediction of ECL values from 2D-FARI is described in Paper VI.

A reference sample containing fatty acids with 2D-FARI values defined in Table 2 in Paper II was analysed with various GC programs and the ECL values were calculated for the compounds in the reference sample. A multivariate calibration was then used to find the relationship between the 2D-FARI (x-variables) and ECL (y-variable) for the FAMEs in the reference sample. The multivariate calibration model was thereafter applied to predict the ECL of other FAMEs from their 2D-FARI values.

Since the 2D-FARI values for the FAMEs in the calibration sample are given by definition, the only information necessary to calculate the ECL value for a compound run under identical conditions as the calibration sample is the 2D-FARI values for the compound, which may be acquired from previously reported data.

Results for prediction of ECLs for various fatty acids with 0–5 double bonds (All data from Paper VI) are shown in Figure 19. The prediction errors of the models based on 2D-FARI values are much smaller than achieved with the models based on the molecular structure (Figure 18). The RMSEP of general models (large range of ECL) based on the 2D-FARI values was less than half the RMSEP for the best models in Paper V. However, it is difficult to make a direct comparison of the two approaches because the predictions were based on different compounds. There are also differences in the validation procedures. Most of the models in Paper V were validated by cross-validation, while the models based on 2D-FARI are validated on test-sets.

The calibration sample contains only saturated and *cis* unsaturated fatty acids. It was therefore a question how accurate the models will predict the ECLs of other types of fatty acids. Because *trans* double bonds show a different behaviour with respect to the ECL shifts required for calculation of 2D-FARI than *cis* double bonds, the method was tested also on *trans* isomers of EPA and DHA. The errors for the predicted ECL values of the *trans* isomers were lower than the errors for the *cis* unsaturated fatty acids.

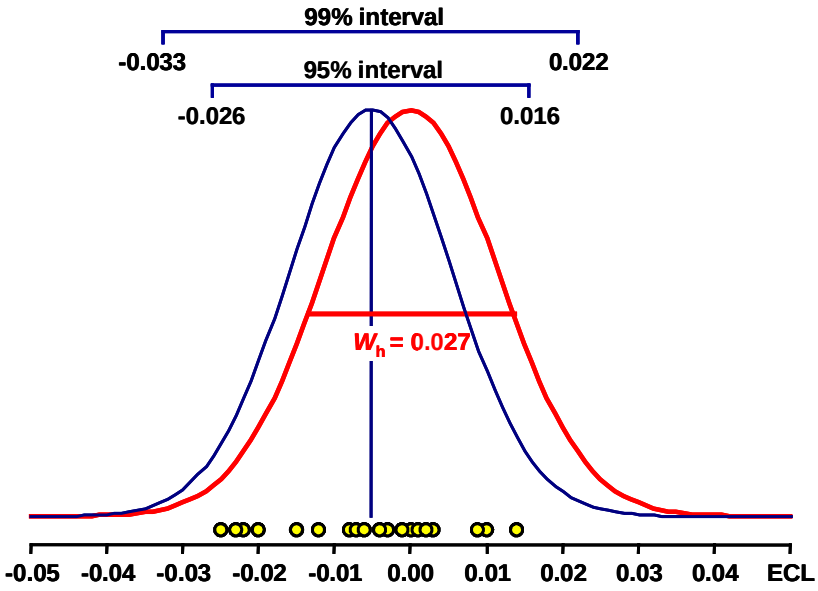



**Figure 19** *Distribution of the residuals (blue curve) for ECL data predicted from 2D-FARI values compared to the chromatographic peak width. The peak width (red curve) is estimated from 20:5 n–3. Residuals of the single fatty acids are shown with yellow dots. The intervals indicate 95 and 99 % of the area of the error distribution curve. RMSEP = 0.012 (Redrawn from Fig. 1a in Paper VI)*

# 5. Identification of FAME from mass spectra

This chapter is a discussion of Papers VII–VIII, which are about application of multivariate analysis for identification of PUFA from mass spectra of FAME. PCA is applied in Paper VII to distinguish between *cis* and *trans* isomers of PUFA. Multivariate regression (PLSR and MLR) is applied in Paper VIII for the determination of the number of double bonds in PUFA.

## 5.1. Identification of *trans* double bonds in PUFA

*Trans* isomers of fatty acids analysed by gas chromatography are usually identified from retention data or occasionally by application of infrared detectors [126–130]. Several reviews have been published on the topic [131–133].

Little attention has been paid to the possible identification of *trans* geometry in fatty acids by MS. However, it has been shown that principal component analysis can distinguish between small differences in mass spectra of picolinyl esters of geometric isomers of monounsaturated fatty acids [73]. Differences have also been found between CI spectra of *cis* and *trans* FAME isomers [60]. It has been shown that *cis* and *trans* geometry in aliphatic chains can be detected by EI-MS [134], and differences between the spectra of *cis* and *trans* monoenoic FAME have been observed when the double bond is close to the carbonyl group [135,136].

The analysis of *cis* and *trans* isomers of PUFA with MI double bond systems in Paper VII showed surprisingly large discrimination caused by the geometries of the double bonds. A PCA score plot of the mass spectra of geometrical isomers in 18:3 *n*–3 is shown in Figure 20. The main variance in the dataset, 77% explained by PC1, is caused by differences in the geometry of the central double bond. There is also a clear discrimination along PC2, which is caused by differences in the terminal double bond. There was no clear discrimination caused by the geometry of the Δ9 double bond.

The effect of *trans* geometry in the central double bond in the triene system can be seen in Figure 21 where the all-*cis* 18:3 *n*–3 is compared to the isomer with *trans* geometry in the Δ12 position. The most notable difference is that there is a substantial reduction of the *m/z* 79 ion, which is normally the base peak in PUFA with three or more MI double bonds. Furthermore, the diagnostic ions for the position of the double bond system [38–40] (α and ω ions, Figure 5d) are reduced and the α-ion is not visible above the background. Similar results were also achieved with 18:3 *n*–6, 20:3 *n*–3 and 22:3 *n*–3. Isomers of 20:3 and 22:3 that

differed only in the geometry of the double bond nearest the carbonyl group were also separated by PCA (Paper VII).

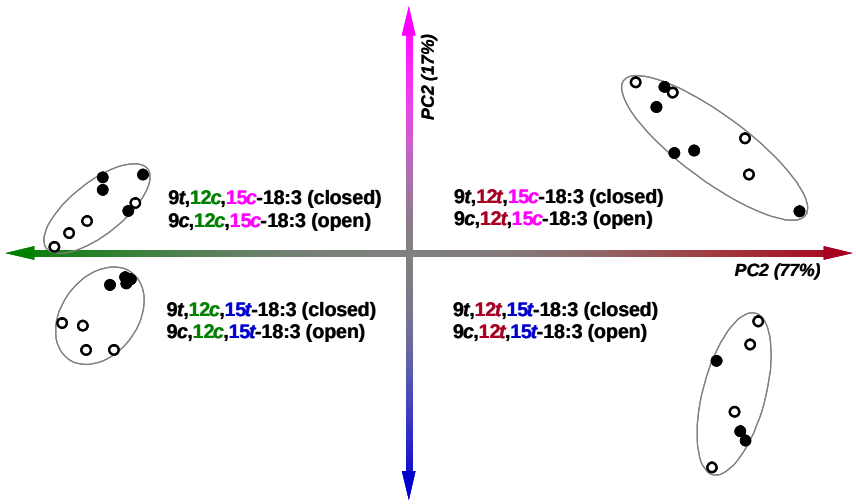



**Figure 20** *PCA score plot of geometrical 18:3 n–3 isomers (Reproduced from Paper VII).*

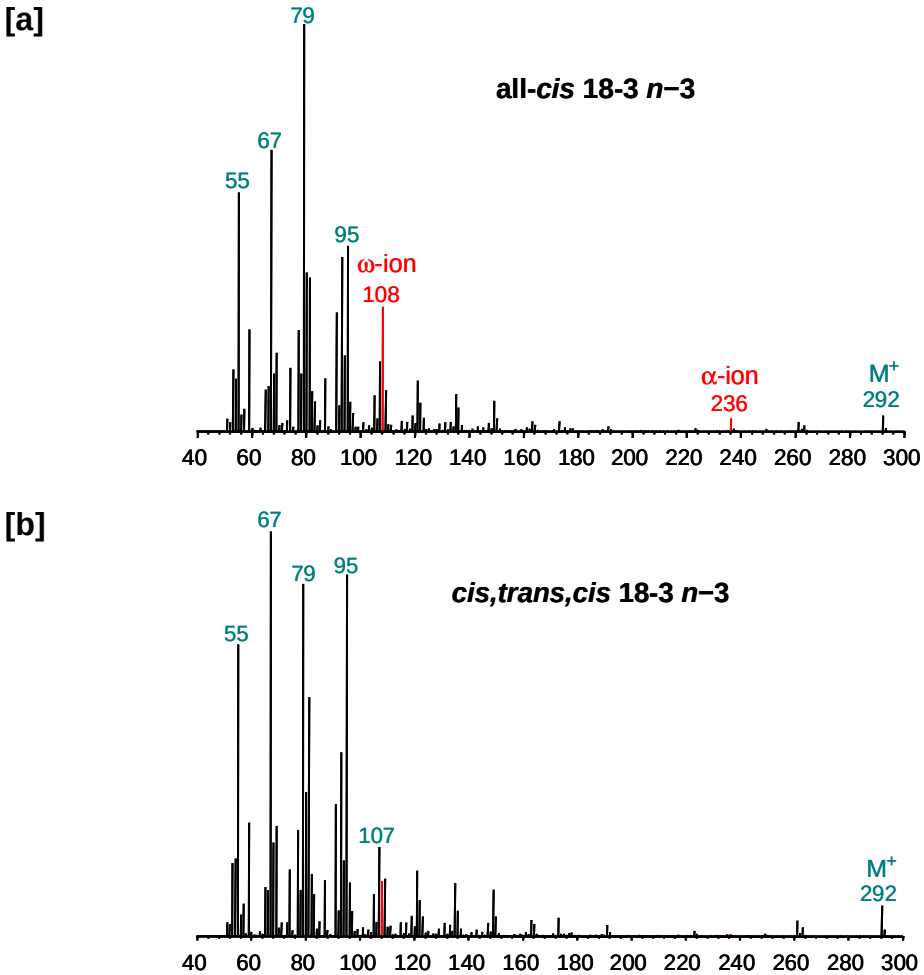



**Figure 21** *PCA score plot of geometrical 18:3 n–3 isomers. (From Paper VII)*

The *trans* isomers of 20:4 *n*–6 and 22:4 *n*–6 were also investigated in Paper VII. Because of the low number of peaks that were completely resolved by GC, conclusive results about the methylene-interrupted tetraenes could not be achieved. However, similar reductions in *m/z* 79 were also observed in tetraenes with *trans* double bonds and observations on 1-*trans* isomers with tentative identifications from the retention times showed the following:

1) *Trans* geometry in the second double bond counted from the methyl end led to a substantial decrease or disappearance of the ω-ion.

2) *Trans* geometry in the second double bond counted from the carbonyl group led to a substantial decrease or disappearance of the α-ion.

3) Absence or low abundance of both the α and ω ions was accompanied by a large reduction in *m/z* 79.

4) Absence or low abundance of one of these ions was accompanied by less reduction in *m/z* 79.

More unsaturated fatty acids was not investigated in Paper VII, but *trans* isomers of 20:5 *n*–3 and 22:6 *n*–3 were analysed in full-scan mode in connection with the work described in Paper III. It is known from the analysis of trienes that *trans* geometry in the *n*–3 and *n*–6 positions lead to a substantial reduction in the retention time, while *trans* geometry in the other positions have little effect or may give slightly higher retention times [137–139, Paper I-VII]. The 1-*trans* isomers of EPA and DHA with lowest retention times are therefore expected to be the isomers with *trans* geometry in the *n*–3 position and the second peaks is expected to be the isomers with *trans* geometry in the *n*–6 position. Thus, the two first 1-*trans* isomers of EPA shown in Figure 17e are expected to be Δ17 *trans* EPA and Δ14 *trans* EPA, and the two first isomers of 1-*trans* DHA in Figure 17f are expected to be Δ19 *trans* DHA and Δ16 *trans* DHA.

According to the results from Paper VII *trans* geometry in the *n*–6 position should give a significant reduction in *m/z* 108 and similar reductions should also be found in the all-*trans* isomers. Isomers with *trans* geometry in the *n*–3 positions should be more similar to the all-*cis* isomers.

The expected reductions were observed for the all-*trans* isomers. Relative to *m/z* 79 (base peak in most isomers), *m/z* 108 was 18 % in all-*cis* EPA and 20 % in all-*cis* DHA, and only 6 % in both all-*trans* isomers. Low abundance in *m/z* 108 was also observed for the isomers with *trans* geometry in the *n*–6 double bonds, 5.3 % for EPA and 5.6 % for DHA. As expected, the levels in the isomers with *trans*

geometry in the *n*–3 position was similar to the all-*cis* isomers, 19 % for EPA and 22 % for DHA.

Similar reductions in the α-ions, which are *m/z* 180 for EPA and *m/z* 166 for DHA, were also observed in spectra of the all-*trans* isomers. The α-ion was 4.5 % relative to *m/z* 79 in all-*cis* EPA and 1.7 % in all-*trans* EPA. The corresponding values for all-*cis* and all-*trans* DHA were 4.5 % and 1.3 %.

## 5.2. Identification of PUFA from selected ions

With scanning mass spectrometers, like sector instruments and quadrupoles, a substantially increased sensitivity can be achieved by selected ion monitoring (SIM), where a few ions are monitored instead of the entire mass spectrum [140]. However, the selection of a small subset of ions may lead to loss of information required for the identification of the compounds. The purpose of the work described in Paper VIII was to find subsets of ions that contain information suitable for identification of fatty acids. The work focused on the determination of the number of double bonds in PUFA with MI double bond systems. If the number of double bonds is known, the chain length and the position of the double bond system can usually be determined from the chromatographic retention times or ECL values.

As can be seen in Figures 5 and 21, the majority of the MS signals are from ions with masses from *m/z* 50 to *m/z* 110, while important diagnostic ions (α-ion, ω-ion, molecular ion) have high masses and low intensity. This has two important consequences. The low intensities of diagnostic ions mean that identification of a compound requires a high-quality spectrum. Acquiring high-quality spectra of unknown compounds is often difficult, since the unknowns are rarely among the largest peaks in a chromatogram. Another consequence of the low abundance in the higher-mass region is that scanning for diagnostic ions will ‘waste’ signal strength because the mass spectrometer is scanning in regions where there are only ions of low abundance.

### *5.2.1 Survey of scan spectra*

Thirty PUFAs were isolated by silver-ion LC and applied in the study. Two of these had NMI double bond systems. In addition, saturated, monounsaturated and di-unsaturated fatty acids were included in the dataset. A survey of the information in the dataset in the mass region from *m/z* 50 to *m/z* 110 was performed by PCA. The score plots are shown in Figure 22a and b. In Figure 22a it can be seen that PUFA are clearly separated from the other classes, and also that NMI-PUFA are separated from MI-PUFA. It has recently been reported that PCA on selected fragments from CI-MS spectra of FAME with methane as reagent gas gave similar discriminations as shown in Figure 22 [56]. It has also been shown that PCA on EI-MS spectra

gives a clear separation between dienes with conjugated and methylene-interrupted systems [141].

A recalculation of the MI-PUFA group gave the score plot in Figure 22b. The *n*–3, *n*–6 and *n*–4 groups are clearly separated along PC2. The *n*–1 fatty acids are positioned between the *n*–3 and *n*–4 group, but very close to the *n*–3 group. The separation along PC1 seems to be related to the Δ-position of the double bond system. The Δ4 (22:5 *n*–6, 22:6 *n*–3, 16:4 *n*–3) and Δ5 (18:4 *n*–4, 18:5 *n*–1, 20:5 *n*–3, 20:4 *n*–6) PUFAs are all positioned to the right in the score plot and fatty acids with high Δ-positions are positioned to the left.

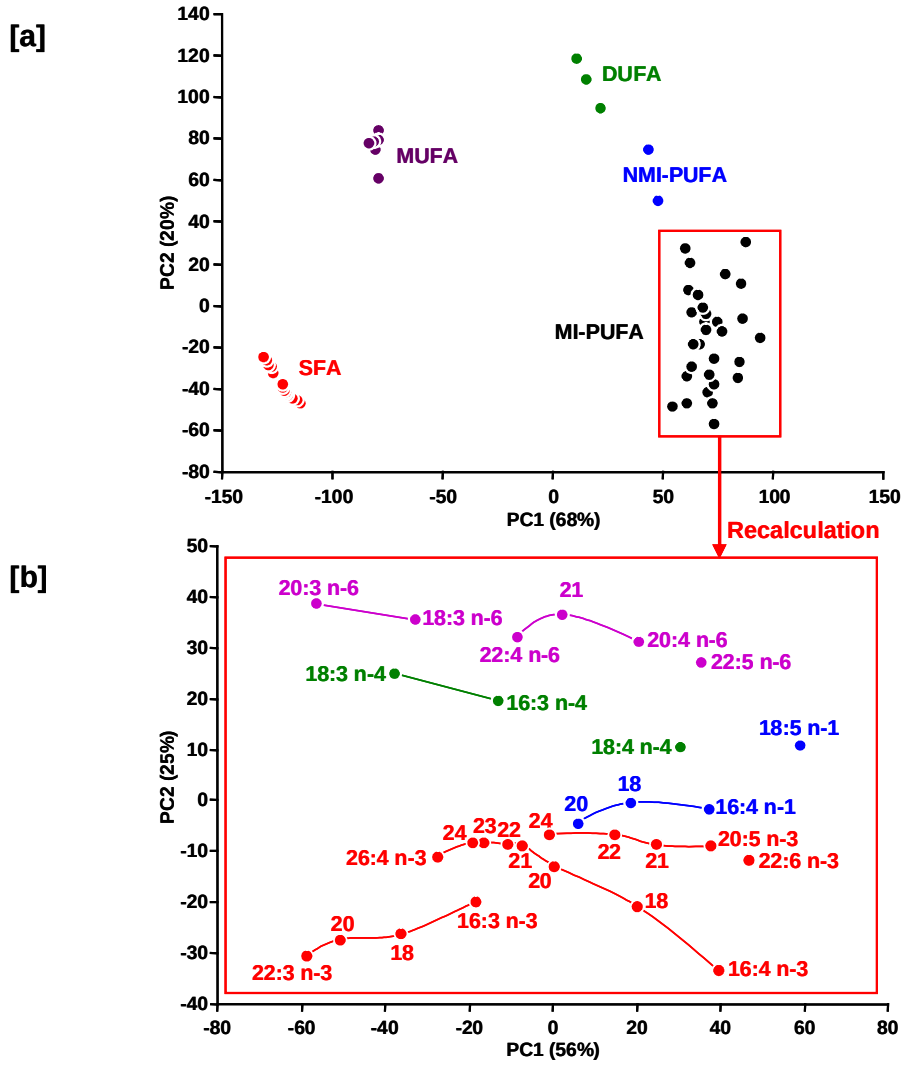



**Figure 22** *PCA score plots of mass spectra (m/z 50–110) of different fatty acid classes [a] and PUFA with methylene-interrupted double bonds [b]. SFA: saturated fatty acids, MUFA: monounsaturated fatty acids, DUFA: diunsaturated fatty acids, NMI-PUFA: polyunsaturated fatty acids with non-methylene interrupted double bond systems, MI-PUFA polyunsaturated fatty acids with methylene interrupted double bond systems. Fatty acids in the same homologous series are connected with curves.*

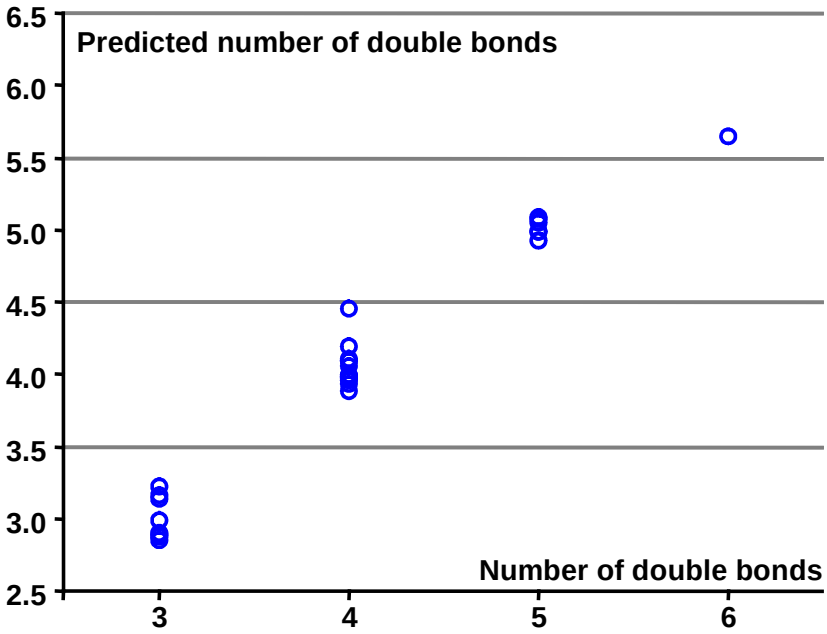



**Figure 23** *Cross validation predictions versus measured values for a PLSR model of the number of double bonds from mass spectra (m/z 50–110) of PUFA. SEP = 0.141. Horizontal lines mark boundaries where the predicted values will be rounded to an incorrect integer. Redrawn from Figure 3 in Paper VIII.*

The two first principal components, explaining 81 % of the variance in the dataset, showed no direct relation to the number of double bonds. However, information about the number of double bonds could be extracted by multivariate regression. PLSR was applied to build a model for the prediction of the number of double bonds based on all ions in the *m/z* 50–110 dataset. The PLSR model had SEP of 0.141 and bias of 0.01. Predicted versus real values are shown in Figure 23. The model will fail in cases where the predicted values are rounded to an incorrect integer, *i.e.* if the prediction error is larger than 0.5. All the predicted values were within these limits. An estimated number of prediction errors (ENPE) for the model can be calculated from the area of the error distribution curve that is outside 0.5. ENPE is only 0.05 % in the case shown in Figure 23.

### *5.2.2. Selection of ions for SIM*

The PLSR on the *m/z* 50–110 spectra shows that there is enough information in the spectral region to find an accurate model for the prediction of the number of double bonds. The optimal number of PLS-components in the PLSR-model gives a rough indication of the minimum number of ions that must be included in a multivariate model without loss of accuracy. The residual y-variance for PLSR-models with different number of PLS-components is shown in Figure 3b in Paper VIII. The residuals falls rapidly from zero to four components, there is a minimum at five

components and a slight increase as more components are included. It was therefore decided to search for suitable SIM combinations by testing MLR-models with three to six ions. There are 61 ions in the dataset with masses from *m/z* 50 to 110 and the number of possible subsets with three, four, five and six ions is 35 990, 521 855, 5 949 147 and 55 523 372.

There are several common techniques for selection of the best combination of variables such as forward selection, backward elimination and stepwise regression. There are also more advanced methods for variable selection, *e.g.* genetic algorithms, which have been applied on MS data [66]. However, there is no guarantee that these methods will find the best possible subset. The optimal subsets of ions were therefore found by testing every combination of three to six ions. To resemble SIM spectra, all subsets were normalised to constant sum prior to the regressions. MLR was used as regression method. To save computation time, regressions were not cross validated in the initial search but the ten models with lowest SEC were subsequently cross-validated. SEP and SEC for these models were highly correlated, indicating that the same subsets would have been found if cross-validation had been applied in the initial step.

SEP for the best subset with six ions (*m/z* 55, 79, 80, 91, 93, 106) gave SEP of 0.132, which was slightly better than SEP for the PLSR on all ions (Figure 23). The subset with *m/z* 55, 74, 79, 80, 91, 93 has been applied in later studies (Papers II, III, IV, V and VI). The selection of this subset is a compromise where the signal strength and separation of other fatty acid classes were also considered. SEP for this subset was 0.148, which gave an ENPE of 0.07 % that was regarded as acceptable.

# 6. Deconvolution of overlapping peaks

A common problem in chromatography is overlapping or partially resolved peaks, which hinders accurate quantification of the overlapping compounds. In cases of severe overlap, it may also be difficult to extract pure spectra or determine the retention time. As shown in Section 1.5 there exist a theoretical framework for deconvolution of overlapping peaks if the detector provides spectra of the analytes, and several resolution techniques have been developed.

Multivariate deconvolution techniques are applied in Papers IX and X. The work in Paper IX is basically about noise and the quality of the GC–MS data. A method for reduction of noise from scanning mass spectrometers is proposed in this paper and applied in Paper X, which shows a practical application of multivariate deconvolution of overlapping geometrical isomers of 18:3 *n*–3.

## 6.1. The scan effect

As shown in Section 1.5, the principles of multivariate deconvolution methods require that the spectrum of an analyte is described by a single vector, $\mathbf{s}^T$ (Equation 17) and the spectrum should therefore be constant during the elution of a peak. Concentration differences in scanning mass spectrometers during a scan lead to overestimation of the last ions relative to the first ions in a scan when the concentration in the detector is increasing (before peak maximum), and overestimation of the first ions relative to the last ions when the concentration in the detector is decreasing (after peak maximum). In GC–MS, narrow peaks and relatively slow scan speeds often lead to large scan-to-scan variations in the spectra. An example of GC–MS SIM mode of a pure fatty acid standard is shown in Figure 24.

This scan-effect can be reduced by using high scan rates. However, this may lead to reduced sensitivity. The effects of increased scan rate in SIM analyses of six ions on the HP-5972 mass spectrometer are illustrated in Figure 25. After each scanned ion, there is a short time interval, the *inter-ion time lag*, where signals are not recorded. There is a similar larger time lag between two scans, the *inter-scan time lag*. These time lags are constant and were estimated to be 14 and 60 ms respectively [Paper IX]. Reduction of the *dwell time*, the time each ion is recorded, leads to loss of signal because the time where no ions are recorded increase relative to the total scan time.

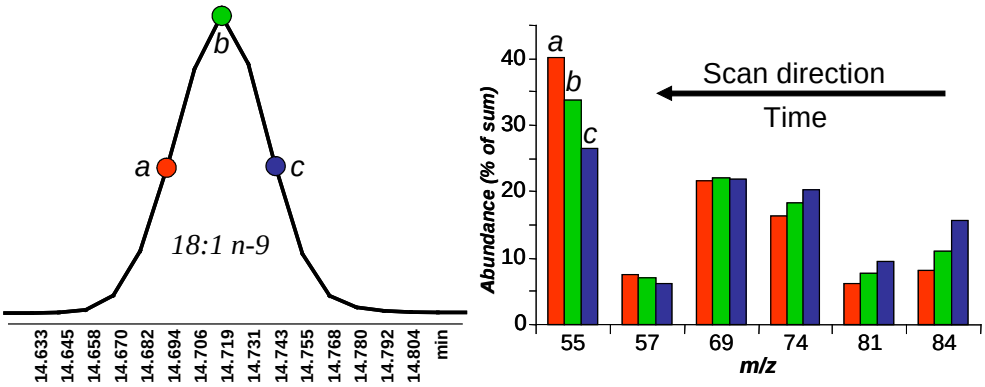



**Figure 24** *Scan effect illustrated by analysis of a pure peak (18:1 n–9 fatty acid methyl ester) analysed with a frequency of 1.36 scans/sec. The bar plot of the masses illustrates how the spectrum changes from the start (a) to the end (c) of the pure peak. The applied instrument (HP-5972, Agilent) scans from the highest to the lowest mass.*

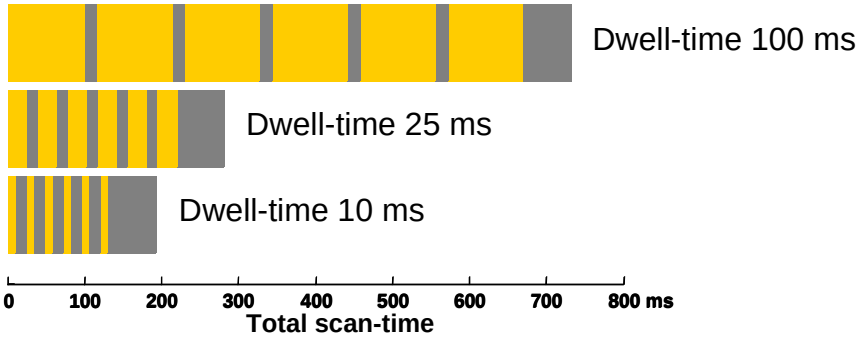



**Figure 25** *Effect of increasing the frequency (scans per second) on the HP-5972 MS detector in SIM mode. Yellow areas represent time when signals are detected; grey areas represent the inter-ion time lags (14 ms) and inter-scan time lags (60 ms) when no ions are recorded. Decreasing the dwell times (the time each ion is recorded) from 100 to 10 ms leads to 60 % reduction of the signal.*

The scan effect can also be reduced by corrections in the chromatographic direction [98,99,142]. The difference in signal abundance, as a function of time for a specific ion in a spectrum, is calculated by interpolation with the signal in the next recorded spectrum. If the time lag between the detection of each ion in the spectrum is known, the signals may be corrected as if all ions were recorded simultaneously. Details are given in the references [98,99,142]. This method has several drawbacks. One is that the shape of the concentration profile cannot be accurately modelled with linear interpolation. It has been proposed to use a second order interpolation from three consecutive spectra to reduce this problem [98].

A second problem is that the correction procedure does not distinguish between differences caused by the scan effect and differences caused by the presence of a second analyte in the spectrum. Consequently, the procedure may correct for differences that are not caused by the scan effect. Details about the data-acquisition must also be known. It is essential to know whether the instrument scans from the highest to the lowest mass or in the reverse direction and corrections for the inter-scan time lag should also be considered.

An alternative solution for reduction of the scan effect is proposed in Paper IX. The method utilizes that mass spectra are non-continuous spectra, as opposed to, for example, UV, IR or NIR spectra. Two consecutive masses in the spectrum are measured almost simultaneously; the scan effect has therefore limited influence on the ratio of these masses. Since a mass spectrum is non-continuous it can be expressed as ratios of consecutive masses without loss of information, but the scan effect is removed or significantly reduced.

Every spectrum is transformed by dividing the intensity of each ion by the intensity of the ion scanned immediately after. The intensity of the last ion recorded in the scan is set to 1. If the inter-scan time lag is low compared to total scan-time it is advisable to divide the abundance of the last ion in the scan by the abundance of the first ion recorded in the next scan. The transformations may also be applied in the reverse direction. After the transformation, each spectrum is normalised to have the same total abundance as the corresponding spectrum in the untransformed data.

These transformations also have drawbacks and limitations. If there are ions in the scan with no signal (except background noise), the noise will be amplified because of division by a noisy signal with abundance near zero. For this reason, these transformations are most feasible for SIM analyses of analytes with similar spectra. The transformations in the spectral direction will radically change the spectra, so they will be less suitable for identification purposes. However, after the resolution, the original spectra, **S,** may be recovered by combining the spectral profiles **C** with untransformed **X** in Equation 19.

A latent projective graph (LPG) [84] is a PCA score plot of the raw data matrix, **X**, with scan numbers (or retention times) as objects and masses (or wave numbers) as variables (See Figure 11). LPG is performed on uncentred data. Noise free pure peaks will be projected as a straight line in LPG because the spectra of a pure peak differ only in the total signal strength. LPGs of two closely eluting 18:1 FAME ($R_s$ = 1.27) are shown in Figure 26. Before ratio-transformation of the spectra (Figure 26a) the peaks appears as two loops because the main variation in the data is the scan-effect and not the difference between the spectra of the two compounds. After transformation (Figure 26b), the peaks are projected as two straight lines because the scan-effect is removed.

Deconvolution of the two monoenes with more severe chromatographic overlap ($R_s$ = 0.68) is shown in Figure 27. Deconvolution of the untransformed **X** failed and gave two profiles with double peak maxima. Deconvolution of the transformed **X** gave to separate peaks with one maximum each. Gentle [95] was applied as deconvolution method in this case because the method does not force unimodality (one maximum) on the chromatographic profiles.

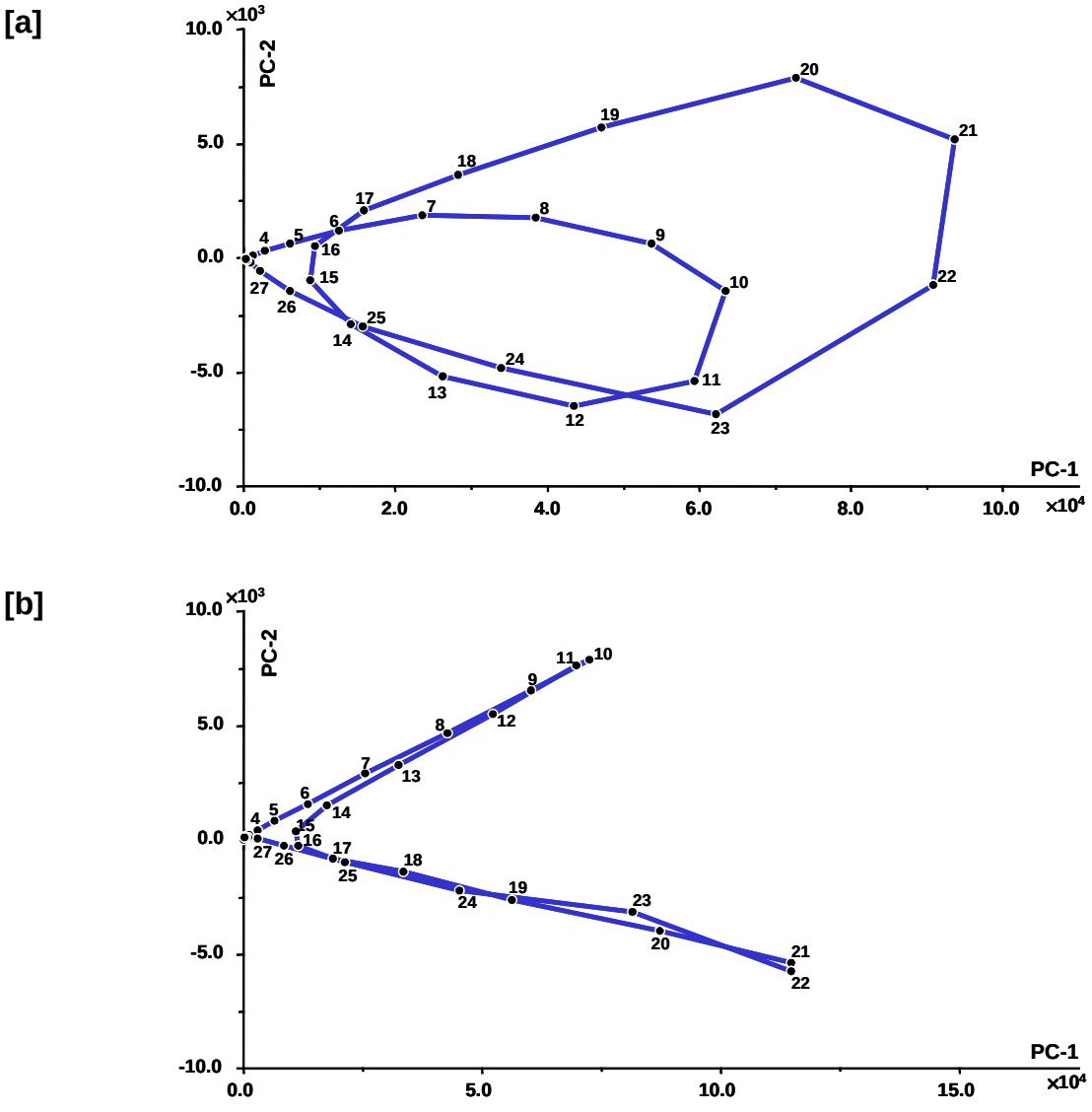



**Figure 26** *Latent projective graphs of 18:1 n–9 and 18:1 n–7 fatty acid methyl esters before [a] and after [b] transformation of the SIM spectra to ratios. Chromatographic resolution, $R_s$, is 1.27 (Reproduced from Paper IX).*

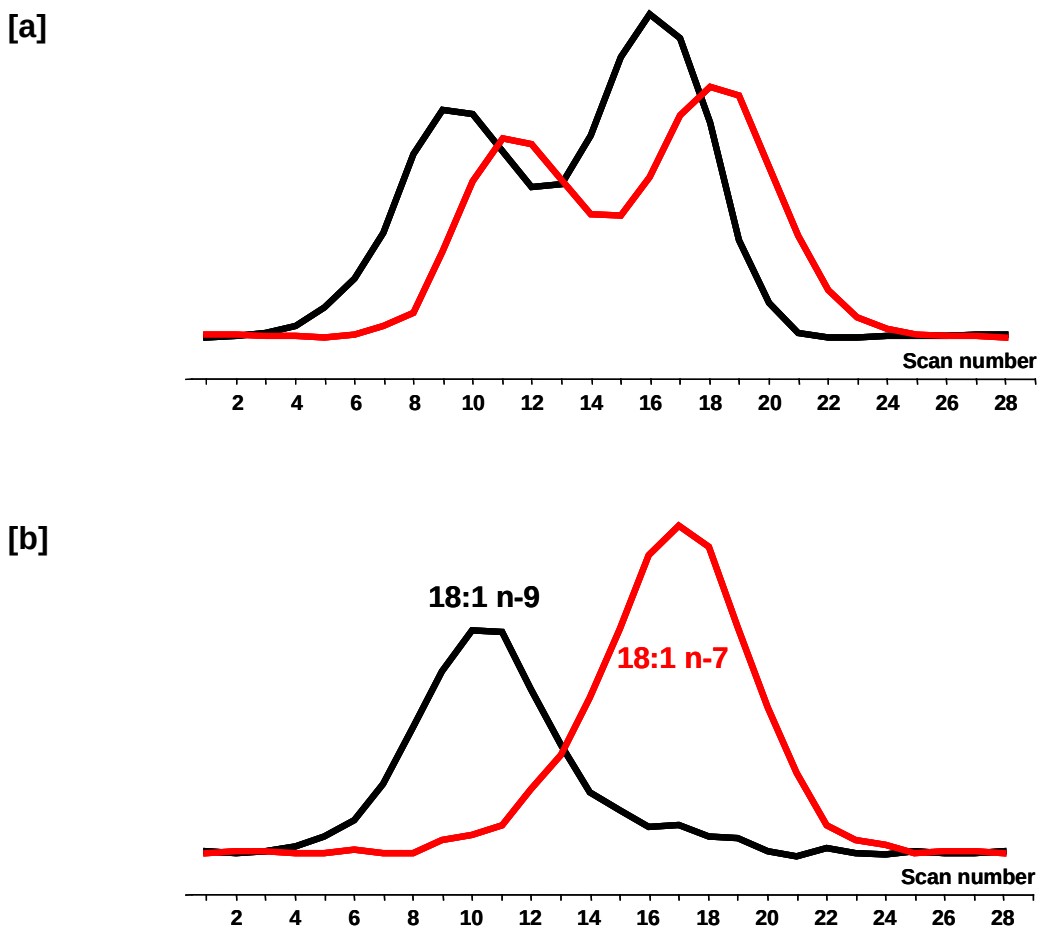



**Figure 27** *The results of multivariate deconvolution by Gentle [95] of 18:1 n–9 and 18:1 n–7 FAMEs before [a] and after [b] transformation of the SIM spectra to ratios. Chromatographic resolution, $R_S$, is 0.68 (Reproduced from Paper IX).*

## 6.2. Multivariate deconvolution of overlapping 18:3 isomers

The spectral transformations proposed in Paper IX were applied in Paper X for resolution of overlapping peaks of 18:3 geometrical isomers.

Because of the similarity in structure and physical properties, complete separation of geometrical 18:3 *n*–3 isomers is hard to achieve. The most difficult resolutions are between the 9*t*,12*t*,15*c* and 9*c*,12*c*,15*t* isomers [143–145, Paper I andVII], and between the 9*c*,12*t*,15*c* and 9*t*,12*c*,15*c* isomers [143,144,146,147,Paper I and VII]. Peak overlaps between 9*c*,12*t*,15*t* and 9*t*,12*c*,15*t* have also been observed [143–146,148]. The quantification of these isomers may be further complicated by overlap with other FAMEs, such as C20 fatty acids [145,148–150] or gamma-linolenic acid (6*c*,9*c*,12*c*-18:3) [145,151,Paper I and VII].

It was shown in Paper VII that the geometry of the central double bond has significant influence on the mass spectra of the isomers. This means that all the resolution problems mentioned above involves two isomers with significant differences in the mass spectra. These differences can be utilised for deconvolution of the overlapping peaks by multivariate methods as outlined in Section 1.5.

The elution patterns of all geometrical isomers of 18:3 *n*–3 and all-*cis* 18:3 *n*–6 with two different temperature programs on BPX-70 are shown in Figure 28. There are five clusters of overlapping isomers, one with three isomers and four with two isomers.

The data were recorded by SIM-MS and the signals for each ion and the deconvoluted chromatographic profiles for clusters a1 and a2 are shown in Figure 29; resolution of the other clusters are shown in Paper X. Overlapping compounds in GC–MS can often be quantified from selective ions, *i.e.* ions that are present in the spectrum in only one of the overlapping compounds. However, in this case there are no selective ions and none of the ions show a clear profile for any of the co-eluting peaks

Since there are no selective ions, no good start-estimates of the chromatographic profiles, **C**, can be found and the resolution must be based on initial estimates of **S** (the purest spectra). Estimates of **S** were found by a modified Borgen procedure according to [92] and the estimates of **C** and **S** were refined by the iterative procedure Gentle [95].

Mixtures of the isomers in various proportions were used to evaluate the quantitative accuracy of the deconvolutions. With few exceptions, $r^2$ was >0.99 for the correlation between predicted and expected areas. The resolution of the cluster with three compounds were less precise, and $r^2$ was in one case as low as 0.86. As can be seen in Figure 29b, there is no region in the chromatogram where the 6*c*,9*c*,12*c* isomer is present without interference of the other compounds. This will lead to poorer start estimates of **S** for cluster a1 than for the other clusters. Further details about the quantitative results are given in Paper X.

The estimation of the correct number of peaks with rank analysis was not considered an issue in this case because all deconvolution are performed on problems where the number of peaks is known. In ‘real’ samples with isomerised linolenic acid, the presence or absence of a certain isomer will normally be known from the presence of other isomers in the sample. However, the number of significant components is a limitation for the number of compounds that can be resolved from one cluster.

Previous studies of the mass spectra of all eight geometric isomers of linolenic acid (Paper VII) indicated that situations as described in Section 1.5.2, where the spectra are identical or linear combinations of other spectra, may occur in peak clusters of linolenic acid isomers. When normalised mass spectra were analysed by PCA, the

two first principal components explained 94% of the variance, and the remaining principal components showed no systematic structure. This indicates that the number of significant factors in a peak cluster with linolenic acid isomers will be maximum three. The number of linolenic acid isomers in a peak cluster should therefore be kept as low as possible, and should not exceed three. The resolution of a peak cluster containing 9*c*,12*t*,15*t*-18:3, 9*t*,12*c*,15*t*-18:3 and 9*t*,12*t*,15*c*-18:3 is shown in Paper IX.

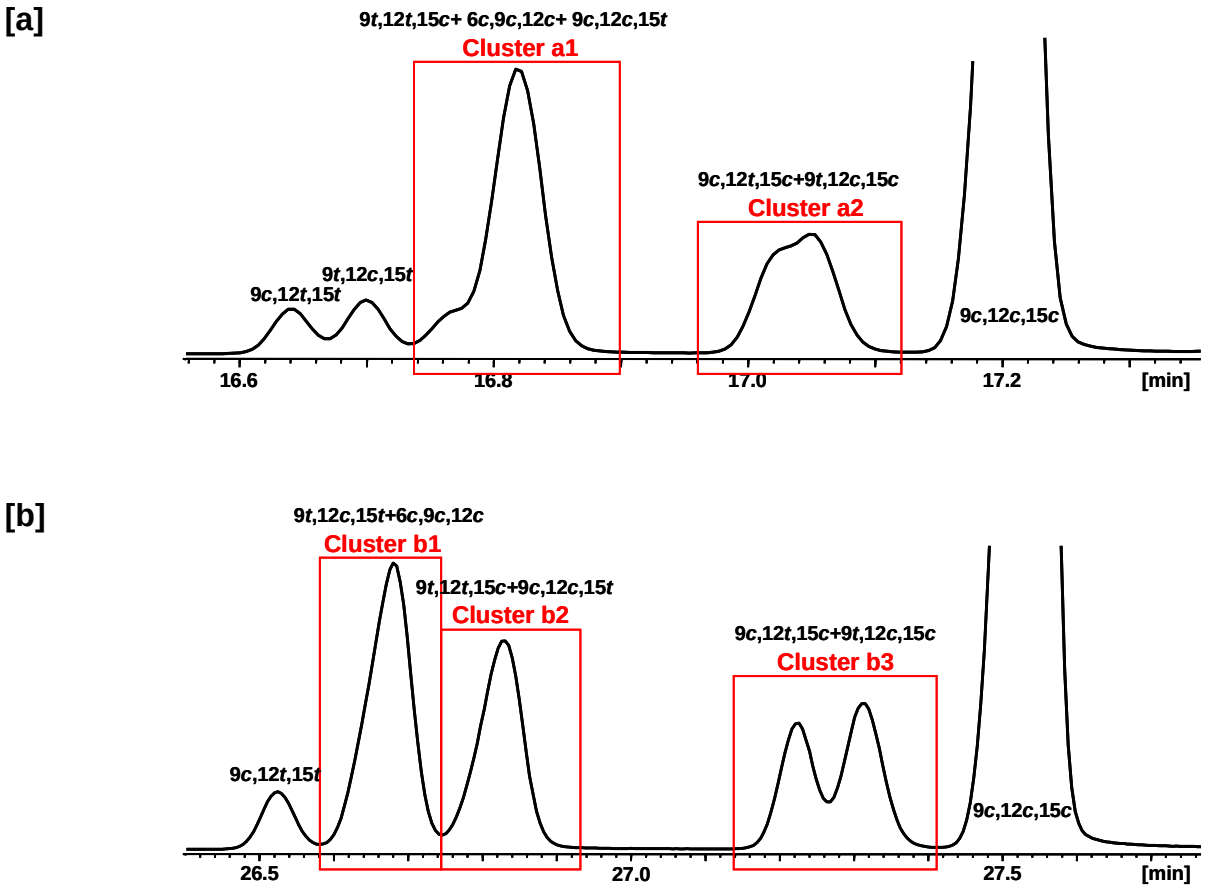



**Figure 28** *Co-elution patters of 18:3 n–3 geometrical isomers with different programs on BPX-70. Additional details are given in Paper X.*

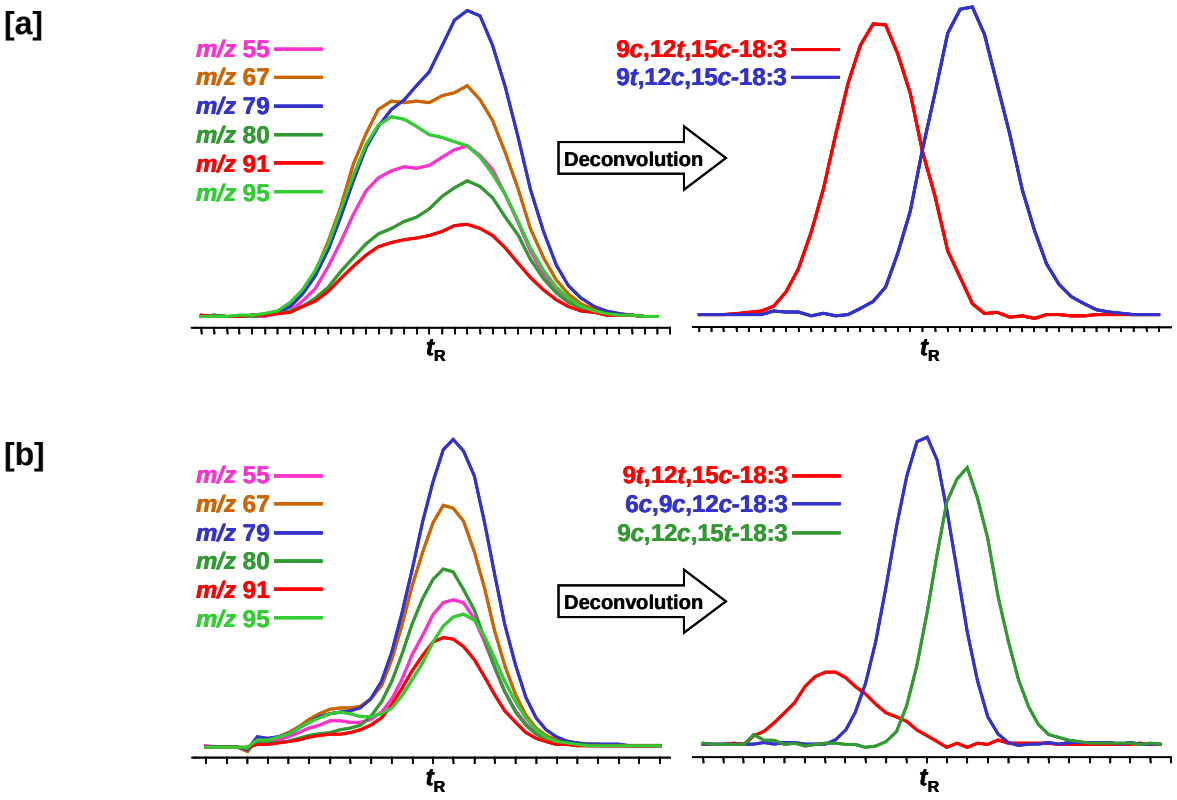



**Figure 29** *Resolution of cluster a2 [a] and a1 [b] in Figure 28. The chromatographic traces of each ion is shown to the left and the deconvoluted profiles to the right.*

# 7. General discussion

The object of this work has not been to devise a general method for fatty acid identifications, but to extend the number of tools available to the analytical chemist working with fatty acids. The various methods may be applied alone or combined, also with existing methodology for fatty acid indentification. Much of the work described in Papers I–VIII provides tools and information that can be applied only on sub-classes of FAME. The strengths, limitations and possible applications of the methods will be briefly discussed in this section.

## 7.1. Limitations and advantages of identification by mathematical models

Identification of compounds based on mathematical models has certain strengths and drawbacks compared to the alternatives. The most important drawback by using multivariate models is that the models are only valid on a similar sub-set of samples as applied in the calibrations. It is therefore essential to know whether the model can be applied on a given spectrum or a set of retention indices. The PLSR models for identification of the number of double bonds in Paper I and VIII can for instance be expected to fail if applied on *trans* fatty acids or fatty acids with NMI double bond systems. To a certain degree, such problems can be overcome by a suitable screening or classification method to reject compounds that cannot be applied in the regression models, like the application of PCA on the spectra in Paper VIII. However, there can be cases where such screening models will fail or cannot be applied.

The main advantage of using mathematical models is that they are highly efficient compared to the alternatives. ‘Manual’ interpretation of spectra and other available information are time consuming, while PCA or regression on a large number of spectra may take fractions of a second. Another advantage is that a mathematical approach may provide more objective estimates of the reliability of the results than manual interpretation of data. Mathematical methods may also be applied in so-called ‘expert systems’ for interpretation of the data that reduce the need of a trained operator [152].

## 7.2. About ‘positive’ and ‘tentative’ identification

All the multivariate models for identification purposes in Papers I, V, VI and VIII provide different types of error estimates for the accuracy or reliability of the predicted result. However, it should be emphasized that the error estimates of the

models are not the same as estimates of the probability that a certain compound is correctly identified. For reasons explained below, estimates for the reliability of identifications can rarely be calculated.

Proposed identifications from the multivariate models should not be regarded as positive identifications of compounds. This is not a problem that is restricted to identifications based on statistics. In many complex systems it may be argued that positive identifications of compounds are not possible and that all identifications should be regarded as tentative. There are numerous examples of incorrect 'identifications' of chemical compounds by chromatographic and spectroscopic techniques. Some examples regarding fatty acids are given in [15].

Especially matching retention times or retention indices of an unknown compound and reference compound should not be regarded as a proof of the identity of the unknown. The risk of false identifications can be reduced by analysing the compounds on two or more columns with different stationary phases, but there is still a risk that compounds with highly similar structures will not be separated.

A similar problem exists with identifications based on mass spectrometry. The mass spectrum of a given compound may be tentatively identified by a match with a reference spectrum in a mass spectral library. This can not be regarded as a proof of the identity since there may exist other compounds with similar mass spectra that are not present in the spectral library.

It is generally not possible to estimate the probability that there are other compounds in a given sample that have identical retention times or similar spectra as the reference. Thus, the probability that a compound is correctly identified cannot be estimated, with the possible exceptions for simple systems (*e.g.* low molecular weight compounds) where the required information about all possible alternatives is available [153].

The identification of a compound is more reliable if it is based on the combination of chromatographic and spectral information. However, one cannot expect that compounds that have very similar mass spectra, *e.g.* geometrical isomers, will be chromatographically separated. The problem of achieving unambiguous identifications, even in some very simple systems, is thoroughly discussed in [154].

## 7.3. Possible applications

### *7.3.1. Use of retention indices*

Papers I and II describe methods for partial identification of fatty acid structure from shifts in ECL values. The methods will basically provide information on the

polarity (number of double bonds) and the chain lengths of the fatty acids. The method is not restricted to any particular class of fatty acids and may also be used to exclude compounds that is not fatty acids, *e.g.* certain hydrocarbons or dimethylacetals that may appear in FAME chromatograms [155,156].

Papers III and IV are restricted to *trans* isomers of 20:5 *n*–3 and 22:6 *n*–3. Although the main purpose of Paper III was to provide data that could be used for optimisation of the elution patterns of the isomers, the paper also provide 2D-FARI data that are suitable for identification of the compounds and equations that can be used to predict the ECL values of the *trans* isomer if the ECL value for the all-*cis* isomer is known.

Papers V and VI describe methods for prediction of ECL values. Paper V is restricted to PUFA with MI double bond systems. Paper VI is not restricted to a particular class of fatty acids, but the 2D-FARI values must be available. Thus, the two methods may be regarded as complementary methods for ECL predictions. There will be cases where only one of the methods can be applied, there will be cases where both methods can be applied, and there will be cases where none on the methods can be applied. Although the focus in these works has been on identification of fatty acids, models that predict ECL values are of value also in method development, because it is possible to foresee chromatographic overlaps of compounds that are not at hand when the chromatographic parameters are optimised.

An advantage of methods based on retention indices is that the reliability of the methods is not dependent on the amounts of the compounds (as long as the peak is detected). In contrast, methods based on spectral information require spectra of a certain quality, which are often not available since unidentified peaks in chromatograms are often small.

### *7.3.2. Mass spectra*

Paper VII and VIII are about multivariate analysis of EI mass spectra of FAME. In Paper VII it was shown that *trans* isomerism in certain positions has significant impact on the mass spectra of PUFA with MI double bond systems. These results are relevant for the work in Paper VIII where it was shown that the number of double bonds in MI-PUFA can be determined from selected ions in the mass spectra. The effects of the *trans* geometry on the mass spectra mean that the models in Paper VIII may not be valid for certain *trans* fatty acids. The results from Paper VII are also a part of the basis for the work in Paper X, where it was shown that the differences in the spectra caused by the *trans* geometry is large enough for deconvolution of overlapping *trans* isomers.

### 7.3.3. Multivariate deconvolutions of chromatographic peaks

Papers IX and X deal with techniques for deconvolution of overlapping FAME peaks. Although the focus in Paper X was basically on the quantification of the isomers, deconvolution of overlapping peaks has relevance also for the identification of the compounds because accurate retention times necessary for calculation of ECL values can be acquired from the resolved chromatographic profiles. Multivariate deconvolutions were applied to find correct retention times of overlapping peaks in Paper II.

The multivariate deconvolution techniques also provide the spectra of the overlapping compounds. It has not been tested whether the deconvoluted spectra are accurate enough to be used for instance in the multivariate regressions in Paper VIII, but deconvoluted spectra will usually be identified by a search against a spectral database (not shown in the papers). The deconvolution techniques and rank estimates (Section 1.5.2 and 4.2) are also efficient tools to verify if chromatographic peaks are pure.

### 7.3.4. Combination of techniques

The various techniques are more informative when combined than when applied separately. As shown in Paper VIII, PCA on the mass spectra may reveal whether a compound is a PUFA with a MI double bond system. If a high-quality full-scan mass spectrum is available, the number of double bonds may be revealed from the molecular ion, which is usually abundant in spectra of FAME with zero to three double bonds. Alternatively, the number of double bonds in MI-PUFA may be determined by regressions as shown in Paper VIII. It was also shown in this paper that PCA gave a class separation according to the position of the double bond system; there is especially a good discrimination between n–3 and n–6 PUFA, which are the two dominating PUFA classes in both terrestrial and marine lipids.

If the total number of double bonds is determined, PCA on retention index shifts (Paper I) or the 2D-FARI method (Paper II) may reveal if there is *trans* geometry in one or several double bonds. Finally, tentative identifications may be verified by comparing the ECL of the unknown peak with expected ECL calculated by the methods proposed in Paper V in the case of MI-PUFA, and Paper VI if 2D-FARI values are available from previous analyses of the same compounds.

It is also possible use library searches on 2D-FARI values from previously identified compounds in the same way as library searches are performed on mass spectra. A library search of 2D-FARI values will give complementary information to library searches of mass spectra. As can be seen by comparing Figure 22b and Figure 15, compounds that have similar mass spectra, *e.g.* members of the same homologous series, are clearly separated in the 2D-FARI plot and compounds with similar 2D-

FARI values (same number of carbons and double bonds, but different in double bond positions) have clearly different mass spectra.

## 7.4. Concluding remarks

There are a number of well established analytical techniques for fatty acid analysis by GC–MS, such as the use of nitrogen containing derivatives of the hydroxyl group, *e.g.* picolinyl [42–44] and DMOX [45–46] derivatives that are applied for determination of the position of double bonds and other functional groups. The number and positions of double bonds can also be determined by derivatisation of the double bonds, *e.g.* hydrogenation with deuterium, silylation [47], or preparation of dimethyl disulfide adducts [48,49]. Pre-fractionations of the samples by LC or TLC are also techniques of importance. The main drawbacks of the above methods are that pre-fractionation of samples, preparation of derivatives and visual interpretation of mass spectra are time consuming operations, and also that some derivatives have poor chromatographic properties.

Even though there are more reliable methods for fatty acid identification, the majority of the fatty acid analyses are still performed with a single gas chromatographic separation of FAME with flame ionization detection (FID) that provides no spectral information. The main reason for this is the simplicity of the FAME preparation and speed of analysis.

Analysis of FAME by electron impact GC–MS can be regarded as a compromise between the use of FID and more thorough but laborious techniques for fatty acid identification. GC–MS of FAME will in general provide less reliable identifications than, for example, GC–MS of picolinyl esters. However, the cost of the analyses will be comparable to GC-FID.

# 8. References


1 Rezanka, T. and Votruba, J. (2002) Chromatography of very long-chain fatty acids from animal and plant kingdoms. *Anal. Chim. Acta* **465**, 273–297.

2 Käkelä, R., Ackman, R.G. and Hyvärinen, H. (1995) Very long chain polyunsaturated fatty acids in the blubber of ringed seals (Phoca hispida sp.) from Laike Saimaa, Lake Ladoga, The Baltic Sea, and Spitsbergen. *Lipids* **30**, 725–731.

3 Anon. (1978) The nomenclature of lipids. *J. Lipid Res.* **19**, 114–128.

4 Spitzer, V. (1999) Screening analyses of unknown seed oils. *Fett/lipid* **101**, 2–19.

5 Brondz, I. (2002) Development of fatty acid analysis by high-performance liquid chromatography, gas chromatography, and related techniques. *Anal. Chim. Acta* **465**, 1–37.

6 Mu, H., Wesén, C. and Sundin, P. (1997) Halogenated fatty acids. 1. Formation and occurrence in lipids. *TrAC, Trends Anal. Chem.* **16**, 266–274.

7 Dembitsky, V.M. and Srebnik, M. (2002) Natural halogenated fatty acids: their analogues and derivatives. *Progr. Lipid Res.* **41**, 315–367.

8 Rosenfeld, J.M. (2002) Application of analytical derivatizations to the quanttitative and qualitative determination of fatty acids. *Anal. Chim. Acta* **465**, 93–100.

9 Ettre, L.S. (1993) Nomenclature for chromatography. *Pure Appl. Chem.* **65**, 819–872.

10 Pap, T.L. and Papai, Z. (2001) Application of a new mathematical function for describing chromatographic peaks. *J. Chromatogr. A* **930**, 53–60.

11 Pápai, Z. and Pap, T.L. (2002) Determination of chromatographic peak parameters by non-linear curve fitting using statistical moments. *Analyst* **127**, 494–498.

12 Pápai, Z. and Pap, T.L. (2002) Analysis of peak asymmetry in chromatography. *J. Chromatogr. A* **953**, 31–38.

13 Ettre, L.S. (2003) More peak asymmetry calculation. *LC-GC Europe* **16**, 192–193.

14 James, A.T. and Martin, J.P. (1952) Gas-liquid partition chromatography: The separation and micro-estimation of volatile fatty acids from formic acid to dodecanoic acid. *Biochem. J.* **50**, 679–690.

15 Ackman, R.G. (2002) The gas chromatograph in practical analyses of common and uncommon fatty acids for the 21st century. *Anal. Chim. Acta* **465**, 175–192.

16 Mayzaud, P. and Ackman, R.G. (1978) The 6,9,12,15,18-Heneicosapentaenoic acid of seal oil. *Lipids* **13**, 24–28.

17 Castello, G., Vezzani, S. and D'Amato, G. (1997) Effect of temperature on the polarity of some stationary phases for gas chromatography. *J. Chromatogr. A* **779**, 275–286.

18 Kováts, E. (1958) Gas-chromatographishe Charakterisierung organischer Verbindungen. Teil 1: Retentionsindices aliphatischer Halogenide, Alkohole, Aldehyde und Ketone. *Helv. Chim. Acta* **41**, 1915–1932.

19 van den Dool, H. and Kratz, P.D. (1963) A generalization of the retention index system including linear temperature programmed gas-liquid partition chromatography. *J. Chromatogr.* **11**, 463–471.

20 Heeg, F.J., Zinburg, R., Neu, H.-J. and Ballschmiter, K. (1979) Berechnung von Retentionsindices auf der Grundlage eines nichtlinearen Zusammenhanges zwischen Nettoretentionzeit und Kolhlenstoffzahl auf n-Alkanen. *Chromatographia* **12**, 451–458.

21 Lebrón-Aguilar, R., Quintanilla-López, J.E. and García-Domínguez, J.A. (2002) Improving the accuracy of Kováts retention indices in isothermal gas chromatography. *J. Chromatogr. A* **945**, 185–194.

22 Gonzales, F.R., Alessandrini, J.L. and Nardillo, A.M. (1998) Considerations on the dependence of gas-liquid chromatographic retention of n-alkanes with the carbon number. *J. Chromatogr. A* **810**, 105–117.

23 García-Domínguez, J.A. and Santiuste, J.M. (1991) Cubic splines compared with other methods for the calculation of programmed temperature retention indices. *Chromatographia* **32**, 116–124.

24 Gonzalez, F.R. and Nardillo, A.M. (1999) Retention index in temperature-programmed gas chromatography. *J. Chromatogr. A* **842**, 29–49.

25 Mjøs, S.A., Meier, S. and Boitsov, S. (2006) Alkylphenol retention indices. *J. Chromatogr. A* **1123**, 98–105.

26 Castello, G. (1999) Retention index systems: Alternatives to the n-alkanes as calibration standards. *J. Chromatogr. A* **842**, 51–64.

27 Evans, M.B. (1989) Recent developments in the gas chromatographic retention index scheme. *J. Chromatogr.* **472**, 93–127.

28 Miwa, T.K., Mikolajczak, K.L., Earle, F.R. and Wolff, I.A. (1960) Gas chromatographic characterization of fatty acids. *Anal. Chem.* **32**, 1739–1742.

29 James, A.T. (1959) Determination of the degree of unsaturation of long chain fatty acids by gas-liquid chromatography. *J. Chromatogr.* **2**, 552–561.

30 Ackman, R.G. (1963) Influence of column temperature in the gas-liquid chromatographic separation of methyl esters of fatty acids on polyester substrates. *J. Gas Chromatogr.* **1**, 11–16.

31 Ackman, R.G. (1963) Structural correlation of unsaturated fatty acid esters through graphical comparison of gas-liquid chromatographic retention times on a polyester substrate. *J. Am. Oil Chem. Soc.* **40**, 558–564.

32 Barve, J.A., Gunstone, F.D., Jacobsberg, F.R. and Winlow, P. (1972) Behaviour of all the methyl octadecenoates and octadecynoates in argentation chromatography and gas-liquid chromatography. *Chem. Phys. Lipids* **8**, 117–126.

33 Ackman, R.G. and Hooper, S.N. (1973) Addiditivity of retention data for ethylenic functions in aliphatic fatty acids. I. Apiezon L. *J. Chromatogr.* **86**, 73–81.

34 Ackman, R.G., Manzer, A. and Joseph, J. (1974) Tentative identification of an unusual naturally-occuring polyenoic fatty acid by calculations from precision open-tubular GLC and structural element retention data. *Chromatographia* **7**, 107–114.

35 Sebedio, J.-L. and Ackman, R.G. (1982) Calculation of GLC retention data for some accessible $C_{20}$ isomeric cis-unsaturated fatty acids. *J. Chromatogr. Sci.* **20**, 231–234.

36 Odham, G. and Stenhagen, E. (1972) Fatty acids. In: Waller, G.R., (Ed.) *Biochemical Applications of Mass Spectrometry.* pp. 211–228. N.Y. (U.S.A.): Wiley.

37 Christie, W.W. (1989) Gas Chromatography – Mass spectrometry and fatty acids. In: Christie, W.W., (Ed.) *Gas Chromatography and Lipids,* pp. 161–184. Ayr, Scotland: The Oily Press

38 Holman, R.T. and Rahm, J.J. (1966) Analysis and characterization of polyunsaturated fatty acids. *Prog. Fats Lipids* **9**, 13–90.

39 Brauner, A., Budziewicz, H. and Boland, W. (1982) Studies in chemical ionization mass spectrometry. 5. localization of homoconjugated triene and tetraene units in aliphatic-compounds. *Org. Mass Spectrom.* **17**, 161–164.

40 Fellenberg, A.J., Johnson, D.W., Poulus, A. and Sharp, P. (1987) Simple mass spectrometric differentiation of the n-3, n-6 and n-9 series of methylene interrupted polyenoic acids. *Biomed. Environ. Mass Spectrom.* **14**, 127–130.

41 Christie, W.W. *The Lipid Library* (www.lipidlibrary.co.uk)

42 Harvey, D.J. (1992) Mass spectrometry of picolinyl and other nitrogen-containing derivatives of fatty acids. In: Christie, W.W., (Ed.) *Advances in Lipid Methodology – one,* pp. 19–80. Ayr (Scotland): The Oily Press.

43 Christie, W.W., Brechany, E.Y., Johnson, S.B. and Holman, R. (1986) A comparison of pyrollidide and picolinyl ester derivatives for the identification of fatty acids in natural samples by gas chromatography-mass spectrometry. *Lipids* **21**, 657–661.

44 Harvey, D.J. (1984) Picolinyl derivatives for the structural determination of fatty acids by mass spectrometry: Applications to polyenoic acids, hydroxy acids, di-acids and related compounds. *Biomed. Mass Spectrom.* **11**, 340–347.

45 Zhang, J.Y., Yu, Q.T., Liu, B.N. and Huang, Z.H. (1988) Chemical modification in mass spectrometry IV – 2-alkenyl-4,4-dimethyloxazolines as derivatives for the double bond location of long-chain olefinic acids. *Biomed. Environ. Mass Spectrom.* **15**, 33–44.

46 Spitzer, V. (1997) Structure analysis of fatty acids by gas chromatography – low resolution electron impact mass spectrometry of their 4,4-dimethyloxazoline derivatives. *Prog. Lipid Res.* **35**, 387–408.

47 Choi, M.H. and Chung, B.C. (2000) Diagnostic fragmentation of saturated and unsaturated fatty acids by gas chromatography-mass spectrometry with pentafluorophenyldimethylsilyl derivatization. *Anal. Biochem.* **277**, 271–273.

48 Tanaka, T., Shibata, K., Hino, H., Murashita, T., Kayama, M. and Satouchi, K.

(1997) Purification and gas chromatographic-mass spectrometric characterization of non-methylene interrupted fatty acids incorporated in rat liver. *J. Chromatogr. B* **700**, 1–8.

49 Scribe, O., Guezennec, J., Dagaut, J., Pepe, C. and Saliot, A. (1988) Identification of the position and the stereochemistry of the double bond in monounsaturated fatty acid methyl esters by gas chromatography/mass spectrometry of dimethyl disulfide derivatives. *Anal. Chem.* **60**, 928–931.

50 Sato, D., Ando, Y., Tsujimoto, R. and Kawasaki, K.-I. (2001) Identification of novel nonmethylene-interrupted fatty acids, 7E13E-20:2, 7E,13E,17Z-20:3, 9E,15E,19Z-22:3, and 4Z,9E15E19Z-22:4 in ophiuroidea (brittle star) lipids. *Lipids* **36**, 1371-1375.

51 Imbs, A.B. and Rodkina, S. (2005) Trans positional ethylenic bonds in two dominant isomers of eicosapentaenoic acid from the freshwater sponge Baicalospongia bacillifera. *Lipids* **40**, 963–968.

52 Schmitz, B. and Klein, R.A. (1986) Mass spectrometric localization of carbon-carbon double bonds: A critical review. *Chem. Phys. Lipids* **39**, 285–311.

53 Minnikin, D.E. (1978) Location of double bonds and cyclopropane rings in fatty acids by mass spectrometry. *Chem. Phys. Lipids* **21**, 313–347.

54 Christie, W.W. (1998) Gas chromatography-mass spectrometry methods for structural analysis of fatty acids. *Lipids* **33**, 343–353.

55 Dodds, E.D., McCoy, M.R., Rea, L.D. and Kennish, J.M. (2005) Proton transfer chemical ionization mass spectrometry of fatty acid methyl esters separated by gas chromatography: quantitative aspects. *Eur. J. Lipid Sci. Technol.* **107**, 560–564.

56 Dayhuff, E. and Wells, M.J.M. (2005) Identification of fatty acids in fishes collected from the Ohio River using gas chromatography-mass spectrometry in chemical ionization and electron impact modes. *J. Chromatogr. A.* **1098**, 144–149.

57 Plattner, R.D., Gardner, H.W. and Kleiman, R. (1983) Chemical ionization mass spectrometry of fatty acids: The effect of functional groups on the CI spectra. *J. Am. Oil Chem. Soc.* **60**, 1298–1303.

58 van Pelt, C.K., Carpenter, B.K. and Brenna, J.T. (1999) Studies of structure and mechanism in acetonitrile chemical ionization tandem mass spectrometry of polyunsaturated fatty acids methyl esters. *J. Am. Mass Spectrom.* **10**, 1253–1262.

59 Michaud, A.L., Yurawecz, M.P., Delmonte, P., Corl, B.A., Bauman, D.E. and Brenna, J.T. (2003) Identification and characterization of conjugated fatty acid methyl esters of mixed double bond geometry by acetonitrile chemical ionization tandem mass spectrometry. *Anal. Chem.* **75**, 4925–4930.

60 van Pelt, C.K. and Brenna, T. (1999) Acetonitrile chemical ionization tandem mass spectrometry to locate double bonds in polyunsaturated fatty acid methyl esters. *Anal. Chem.* **71**, 1981–1989.

61 Bro, R., van den Berg, F., Thybo, A., Andersen, C.M., Jørgensen, B.M. and Andersen, H. (2002) Multivariate data analysis as a tool in advanced quality monitoring in the food production chain. *Trends Food Sci. Technol.* **13**, 235–244.

62 Werther, W. and Varmuza, K. (1992) Exploratory data analysis of infrared spectra. *Fres. J. Anal. Chem.* **344**, 223–226.

63 Fang, G. and Liu, N. (2001) Determination of eight essential amino acids in mixtures by chemometrics-spectrophotometry without separation. *Anal. Chim. Acta* **445**, 245–253.

64 Ho, C.N, Christian, G.D. and Davidson, E.R. (1978) *Application of the method of rank annihilation to quantitative analyses of multicomponent fluorescence data from the video fluorometer*. Anal. Chem. **50**, 1108–1113.

65 Demuth, W., Karlovits, M. and Varmuza, K. (2004) Spectral similarity versus structural similarity: mass spectrometry. *Anal. Chim. Acta* **516**, 75–85.

66 Yoshida, H., Leardi, R., Funatsu, K. and Varmuza, K. (2001) Feature selection by genetic algorithms for mass spectral classifiers. *Anal. Chim. Acta* **446**, 485–494.

67 Werther, W., Lohninger, H., Stancl, F. and Varmuza, K. (1994) Classification of mass spectra. A comparison of yes/no classification methods for the recognition of simple structural properties. *Chemom. Int. Lab. Syst.* **22**, 63–74.

68 Scsibrany, H. and Varmuza, K. (1992) Common substructures in groups of compounds exhibiting similar mass spectra. *Fres. J. Anal. Chem.* **344**, 220–222.

69 Brakstad, F. (1995) The feasibility of latent variables applied to GC-MS data. *Chemom. Int. Lab. Syst.* **29**, 157–176.

70 Brakstad, F. (1993) Accurate determination of double bond position in mono-unsaturated straight-chain fatty acid ethyl esters from conventional electron impact mass spectra by quantitative spectrum-structure modelling. *Chemom. Int. Lab. Syst.* **19**, 87–100.

71 Leonhardt, B.A., DeVilbiss, E.D. and Klun, J.A. (1983) Gas chromatographic mass spectrometric indication of double bond position in monounsaturated primary acetates and alcohols without derivatization. *Org. Mass Spectrom.* **18**, 9–11.

72 Lanne, B.S., Appelgrn, M. and Bergström, G. (1985) Determination of the double bond position in monounsaturated acetates from their mass spectra. *Anal. Chem.* **57**, 1621–1625.

73 Leth, T. (1997) Chemometric analysis of mass spectra of cis and trans fatty acid picolinyl esters. *Z. Lebensm. Unters. Forsch. A* **205**, 111–115.

74 Christie, O.H. (1995) Introduction to multivariate methodology, an alternative way? *Chemom. Int. Lab. Syst.* **29**, 177–188.

75 Malinowski, E.N. (2002) *Factor Analysis in Chemistry*, 3rd edn. NY (USA): Wiley.

76 Wold, S., Esbensen, K. and Geladi, P. (1987) Principal Component Analysis. *Chemom. Int. Lab. Syst.* **2**, 37–52.

77 Martens, H. and Næs, T. (1991) *Multivariate Calibration*, NY (USA): Wiley.

78 Manne, R. (1987) Analysis of two partial-least-squares algorithms for multivariate calibration. *Chemom. Int. Lab. Syst.* **2**, 187–197.

79 Kvalheim, O.M. and Karstang, T.V. (1989) Interpretation of latent variable

regression models. *Chemom. Int. Lab. Syst.* **7**, 39–51.

80 Maeder, M. (1987) Evolving factor analysis for the resolution of overlapping chromatographic peaks. *Anal. Chem.* **59**, 527–530.

81 Maeder, M. and Zilian, A. (1988) Evolving factor analysis, a new multivariate technique in chromatography. *Chemom. Int. Lab. Syst.* **3**, 205–213.

82 Keller, H.R. and Massart, D.L. (1991) Peak purity control in liquid chromatography with photodiode-array detection by a fixed size moving window evolving factor analysis. *Anal. Chim. Acta* **246**, 379–390.

83 Toft, J. and Kvalheim, O. (1993) Eigenstructure tracking analysis for revealing noise pattern and lokal rank in instrumental profiles: Application to transmittance and absorbance IR spectroscopy. *Chemom. Int. Lab. Syst.* **19**, 65–73.

84 Kvalheim, O.M. and Liang, Y.-Z. (1992) Heuristic evolving latent projections: Resolving two-way multicomponent data. 1. Selectivity, latent-projective graph, datascope, local rank, and unique resolution. *Anal. Chem.* **64**, 936–946.

85 Liang, Y.-Z. and Kvalheim, O.M. (1994) Diagnosis and resolution of multiwavelength chromatograms by rank map, orthogonal projections and sequential rank analysis. *Anal. Chim. Acta* **292**, 5–15.

86 Grung, B. and Kvalheim, O.M. (1995) Resolution of multicomponent profiles with partial selectivity. A comparison of direct methods. *Chemom. Int. Lab. Syst.* **29**, 75–87.

87 Malinowski, E.R. (1982) Obtaining the key set of typical vectors by factor analysis and subsequent isolation of component spectra. *Anal. Chim. Acta* **134**, 129–137.

88 Schostack, K.J. and Malinowski, E.R. (1989) Preferred set selection by iterative key set factor analysis. *Chemom. Int. Lab. Syst.* **6**, 21–29.

89 Windig, W. and Guilment, J. (1991) Interactive self-modelling mixture analysis. *Anal. Chem.* **63**, 1425–1432.

90 de Juan, A., van den Bogaert, F., Cuesta Sánchez, F. and Massart, D.L. (1996) Application of the needle algorithm for exploratory analysis and resolution of HPLC–DAD data. *Chemom. Int. Lab. Syst.* **33**, 133–145.

91 Cuesta Sánchez, F., Toft, J., van den Bogaert, B. and Massart, D.L. (1996) Orthogonal projection approach applied to peak purity assessment. *Anal. Chem.* **68**, 79–85.

92 Grande, B.-V. and Manne, R. (2000) Use of convexity for finding pure variables in two-way data from mixtures. *Chemom. Int. Lab. Syst.* **50**, 19–33.

93 Karjalainen, E. (1989) The spectrum reconstruction problem. Use of alternating regression for unexpected spectral components in two-dimensional spectroscopies. *Chemom. Int. Lab. Syst.* **7**, 31–38.

94 Vandeginste, B.G.M., Derks, W. and Kateman, G. (1985) Multicomponent self-modelling curve resolution in the high-performance liquid chromatography by iterative target transformation analysis. *Anal. Chim. Acta* **173**, 253–264.

95 Manne, R. and Grande, B.-V. (2000) Resolution of two-way data from hyphenated chromatography by means of elementary matrix transformations.

*Chemom. Int. Lab. Syst.* **50**, 35–46.

96 Tauler, R., Smilde, A. and Kowalski, B. (1995) Selectivity, local rank, three-way data analysis and ambiguity in multivariate curve analysis. *J. Chemom.* **9**, 31–58.

97 Peré-Trepat, E., Lacorte, S. and Tauler, R. (2005) Solving liquid chromatography mass spectrometry coelution problems in the analysis of environmental samples by multivariate curve resolutiuon. *J. Chromatogr. A* **1096**, 111–122.

98 Keller, H.R., Massart, D.L., Kiechle, P. and Erni, F. (1992) Effect on scan time on the methods based on evolving factor analysis in liquid chromatography. *Anal. Chim. Acta* **256**, 125–131.

99 Keller, H.R. and Massart, D.L. (1992) Artefacts in evolving factor analysis-based methods for peak purity control in liquid chromatography with diode-array detection. *Anal. Chim. Acta* **263**, 21–28.

100 Krupcik, J. and Bohov, P. (1985) Use of equivalent chain lengths for the characterization of fatty acid methyl esters separated by linear temperature-programmed gas chromatography. *J. Chromatogr.* **346**, 33–42.

101 Thompson, R.H. (1997) Direct measurement of total trans- and cis-octadecenoic fatty acids based on a gas-liquid chromatographic class separation of trans-18:1 and cis-18:1 fatty acid methyl esters. *J. Chromatogr. Sci.* **35**, 536–544.

102 Bannon, C.D., Craske, J.D. and Norman, L.M. (1988) Effect of overload of capillary gas-liquid chromatographic column on the equivalent chain lengths of c18 unsaturated fatty acid methyl esters. *J. Chromatogr. A* **447**, 43–52.

103 Western, R.J., Lau, S.S.G., Marriot, P.J. and Nichols, P.D. (2002) Positional and geometric isomer separation of FAME by comprehensive 2-D GC. *Lipids* **37**, 715–724.

104 Hénon, G., Kemény, Z., Recseg, K., Zwobada, F. and Kovari, K. (1999) Deodorization of vegetable oils. Part I: Modelling the geometrical isomerization of polyunsaturated fatty acids. *J. Am. Oil Chem. Soc.* **75**, 73–81.

105 Kemény, Z., Recseg, K., Hénon, G., Kóvari, K. and Zwobada, F. (2001) Deodorization of vegetable oils: Prediction of trans polyunsaturated fatty acid content. *J. Am. Oil Chem. Soc.* **78**, 973–979.

106 Wijesundera, R.C., Ratanayake, W.M.N. and Ackman, R.G. (1989) Eicosapentaenoic acid geometrical isomer artifacts in heated fish oil esters. *J. Am. Oil Chem. Soc.* **66**, 1822–1830.

107 Fournier, V., Destaillats, F., Juanéda, P., Dionisi, F., Lambelet, P., Sébédio, J.L. and Berdeux, O. (2006) Thermal degradation of long-chain polyunsaturated fatty acids during deodorization of fish oils. *Eur. J. Lipi Sci. Technol.* **108**, 33–42.

108 Chardigny, J.-M., Sebedio, J.-L., Grandgirard, A., Martine, L., Berdeaux, O. and Vatele, J.-M. (1996) Identification of novel trans isomers of 20:5n-3 in liver lipids of rats fed a heated oil. *Lipids* **31**, 165–168.

109 Chardigny, J.-M., Blond, J.-P., Bretillon, L., Mager, E., Poullain, D., Martine, L., Vatèle, J.-M., Noël, J.-P. and Sébédio, J.-L. (1997) Conversion of 18:3 d9 cis, 12 cis, 15 trans in rat liver microsomes. *Lipids* **32**, 731–735.

110 Mjøs, S.A. and Solvang, M. (2006) Geometrical isomerisation of EPA and DHA

at high temperatures. *Eur. J. Lipid Sci. Technol.* **108**, 589–597.

111 Poole, C.F., Kollie, T.O. and Poole, S.K. (1992) Recent advances in solvation models for stationary phase characterization and the prediction of retention in gas chromatography. *Chromatographia* **34**, 281–302.

112 Abraham, M.H., Whiting, G.S., Doherty, R.M. and Shuely, W. (1990) Hydrogen bonding. XV. A new characterisation of the McReynolds 77-stationary phase set. *J. Chromatogr.* **518**, 329–348.

113 Rorschneider, L. (1965) Die Vorausbereichnung von Gaschromatographischen Retentionzeiten aus Statistisch Ermittelten “Polaritäten”. *J. Chromatogr.* **17**, 1–12.

114 Rorschneider, L. (1966) Eine Methode zur Charakterisierung von gaschromatographischen Trennflüssigkeiten. *J. Chromatogr.* **22**, 6–22.

115 McReynolds, W.O. (1970) Characterization of some liquid phases. *J. Chromatogr. Sci.* **8**, 685–691.

116 Liang, X., Wang, W., Wu, W., Schramm, K.W., Henkelmann, B. and Kettrup, A. (2000) Quantitative relationships between chromatographic retentions and molecular structures of polychlorinated dibenzo-p-dioxins (PCDDs). *Chemosphere* **41**, 923–929.

117 Rayne, S. and Ikonmomu, M.G. (2003) Predicting gas chromatographic retention times for the 209 polybrominated diphenyl ether congeners. *J. Chromatogr. A* **1016**, 235–248.

118 Yin, C., Liu, W., Li, Z., Pan, Z., Lin, T. and Zhang, M. (2001) Chemometrics to chemical modelling: Structural coding in hydrocarbons and retention indices of gas chromatography. *J. Sep. Sci.* **24**, 213–220.

119 Garkani-Nejad, Z., Karlovits, M., Demuth, W., Stimpfl, T., Vycudilik, W., Jalali-Heravi, M. and Varmuza, K. (2004) Prediction of gas chromatographic retention indices of a diverse set of toxicologically relevant compounds. *J. Chromatogr. A* **1028**, 287–295.

120 Peng, C.T., Ding, S.F., Hua, R.L. and Yang, Z.C. (1988) Prediction of retention indexes. I. Structure retention index relationship on apolar columns. *J. Chromatogr.* **436**, 137–172.

121 Peng, C.T., Yang, Z.C. and Ding, S.F. (1991) Prediction of retention indices II. Structure-retention relationships on polar columns. *J. Chromatogr.* **586**, 85–112.

122 Zenkevich, I.G., Moeder, M., Koeller, G. and Schrader, S. (2004) Using new structurally related additive schemes in the precalculation of gas chromatographic retention indices of polychlorinated hydroxybiphenyls on HP-5 stationary phase. *J. Chromatogr. A* **1025**, 227–236.

123 Wijesundera, R.C. and Ackman, R.G. (1989) Evaluation of calculation of ECL values for cis and trans isomers of some diethylenic C20 fatty acids: Mono- and diethylenic capillary GLC data for the liquid phases SP-2340, Supelcowax-10, and SPB-1. *J. Chromatogr. Sci.* **27**, 399–404.

124 Christie, W.W. (1988) Equivalent chain-lengths of methyl ester derivatives of fatty acids on Gas chromatography. *J. Chromatogr.* **447**, 305–314.

125 Scholfield, C.R. and Dutton, H.J. (1970) Gas chromatographic equivalent chain lengths of isomeric methyl octadecenoates and octadecynoates. *J. Am. Oil Chem. Soc.* **47**, 1–2.

126 Mossoba, M.M., McDonald, R.E. and Prosser, A.R. (1993) Gas chromatographic/matrix isolation/fourier transform infrared spectroscopic determination of trans-monounsaturated and saturated fatty acid methyl esters in partially hydrogenated menhaden oil. *J. Agric. Food Chem.* **41**, 1998–2002.

127 Mossoba, M.M. (1993) Applications of capillary GC–FTIR. *Inform* **4**, 854–859.

128 Wahl, H.G., Habel, S.-Y., Schmieder, N. and Liebich, H.M. (1994) Identification of cis/trans isomers of methyl esters and oxazoline derivatives of unsaturated fatty acids using GC-FTIR-MS. *J. High Resol. Chromatogr.* **17**, 543–548.

129 Mossoba, M.M., McDonald, R.E., Roach, J.A.G., Fingerhut, D.D., Yurawecz, M.P. and Sehat, N. (1997) Spectral conformation of *trans* monounsaturated $C_{18}$ fatty acid positional isomers. *J. Am. Oil Chem. Soc.* **74**, 125–130.

130 Sémon, E., Ferary, S., Auger, J. and Le Quéré, J.L. (1998) Gas chromatography–fourier transform infrared spectrometry of fatty acids: New applications with a direct deposition interface. *J. Am. Oil Chem. Soc.* **75**, 101–105.

131 Firestone, D. and Sheppard, A. (1992) Determination of trans fatty acids. In: Christie, W.W., (Ed.) *Advances in Lipid Methodology – one*, pp. 173–322. Ayr, Scotland: The Oily Press.

132 Ratanayake, W.M.N. (1998) Analysis of trans fatty acids. In: Sébédio, J.L. and Christie, W.W., (Eds.) *Trans fatty acids in human nutrition*, pp. 115–161. Dundee, Scotland: The Oily Press.

133 Ledoux, M., Laloux, L. and Wolff, R.L. (2000) Analytical methods for determination of trans-C18 fatty acid isomers in milk fat. A review. *Analusis* **28**, 402–412.

134 Natalis, P. (1965) Geometrical isomerism and mass spectra. *Mass spectrometry – A NATO advanced study institute on theory, design and applications, Glasgow, August 1964 (ed: R.I.Reed), Academic Press* 379–398.

135 Hallgren, B. (1959) The mass spectra of methyl oleate, methyl linoleate and methyl linolenate. *Acta Chem. Scand.* **13**, 845–847.

136 Ryhage, R., Ställberg-Stenhagen, S. and Stenhagen, E. (1961) Methyl esters of α,β-unsaturated long-chain acids. On the structure of C27-phthienoic acid. *Arkiv för Kemi* **18**, 179–194.

137 Juanèda, P., Sebedio, J.L. and Christie, W.W. (1994) Complete separation of the geometrical isomers of linolenic acid by high performance liquid chromatography with a silver ion column. *J. High Resol. Chromatogr.* **17**, 321–324.

138 Wolff, R.L. (1993) Heat-induced geometrical isomerization of alpha-linolenic acid: Effect of temperature and heating time on the appearance of individual isomers. *J. Am. Oil Chem. Soc.* **70**, 425–430.

139 Wolff, R.L., Nour, M. and Bayard, C.C. (1996) Participation of the cis-12 ethylenic bond to cis-trans isomerization of the cis-9 and cis-15 ethylenic bonds in heated alpha-linolenic acid. *J. Am. Oil Chem. Soc.* **73**, 327–332.

140 Turnhofer, S. and Vetter, W. (2005) A gas chromatography/electron impact ionization-mass spectrometry-selected ion monitoring method for the determining the fatty acid pattern in food after formation of fatty acid methyl esters. *J. Agric. Food Chem.* **53**, 8896–8903.

141 Mjøs, S.A. and Pettersen, J. (2001) *Improved methods for analysis of fatty acid isomers. SSF report series A 156.* Norwegian Herring Oil and Meal Industry's Research Institute, Bergen (Norway).

142 Karjalainen, E.K. and Karjalainen, U.P. (1996) *Data analysis for hyphenated techniques*, Amsterdam (The Netherlands): Elsevier.

143 Grandgirard, A., Julliard, F., Prevost, J. and Sebedio, J.L. (1987) Preparation of geometrical isomers of linolenic acid. *J. Am. Oil Chem. Soc.* **64**, 1434–1440.

144 Wolff, R.L. (1992) Resolution of linolenic acid geometrical isomers by gas-liquid chromatography on a capillary column coated with a 100% cyanopropyl polysiloxane film (CP Sil-88). *J. Chromatogr. Sci.* **30**, 17–22.

145 Ratanayake, V.M.N. and Beare-Rogers, J.L. (1990) Problems of analyzing C18 cis- and trans-fatty acids of margarine on the SP-2340 capillary column. *J. Chromatogr. Sci.* **28**, 633–639.

146 Rakoff, H. and Emken, E. (1982) Synthesis and properties of methyl 9,12,15-octadecatrienoate geometric isomers. *Chem. Phys. Lipids* **31**, 215–225.

147 Wolff, R.L. and Sebedio, J.L. (1991) Geometrical isomers of linolenic acid in low-calorie spreads marketed in France. *J. Am. Oil Chem. Soc.* **68**, 719–725.

148 Wolff, R.L. (1992) Trans-polyunsaturated fatty acids in French edible rapeseed and soybean oils. *J. Am. Oil Chem. Soc.* **69**, 106–110.

149 Wolff, R.L. (1993) Further studies on artificial geometrical isomers of alpha-linolenic acid in edible linolenic acid-containing oils. *J. Am. Oil Chem. Soc.* **70**, 219–224.

150 Wolff, R.L. (1994) Analysis of alpha-linolenic acid geometrical isomers in deodorized oils by capillary gas-liquid chromatography on cyanoalkyl polysiloxane stationary phases: A note of caution. *J. Am. Oil Chem. Soc.* **71**, 907–909.

151 Wolff, R.L. and Sébédio, J.-L. (1994) Characterization of gamma-linolenic acid geometrical isomers in borage oil subjected to heat treatments (deodorization). *J. Am. Oil Chem. Soc.* **71**, 117–126.

152 Peris, M. (2002) Present and future of expert systems in food analysis. *Anal. Chim. Acta* **454**, 1–11.

153 Zenkevich, I.G. (2001) Interpretation of retention indices in gas chromatography for establishing structures of isomeric products of alkylarenes radical chlorination. *Russ. J. Org. Chem.* **37**, 270–280.

154 Zenkevich, I.G. (2001) Comparative charcterization of conditions for unambigous chromatographic identification of organic substances. *J. Anal. Chem.* **56**, 915–924.

155 Kayama, M. and Mankura, M. (1998) Natural oleochemicals in marine fishes. *Inform* **9**, 794–799.

156 Caruso, U (1996) *Simple analysis of plasmalogens in erythrocytes using gas chromatography/mass spectrometry with selected ion monitoring acquisition.* Rapid Commun. Mass Spectrom. **10**, 1283–1285.