Svein A. Mjøs

Norwegian Institute of Fisheries and Aquaculture Research, Bergen, Norway

The prediction of fatty acid structure from selected ions in electron impact mass spectra of fatty acid methyl esters

The possibility of fatty acid identification from ions in the mass region of 50–110 amu in electron impact mass spectra has been studied by application of various multivariate techniques. The focus has mainly been on the prediction of the number of double bonds in methylene-interrupted polyunsaturated fatty acids isolated from marine lipids.

The number of double bonds in these fatty acids could be predicted with high accuracy, both by partial least squares regression on all ions in this region, and by multiple linear regressions on selected subsets of 5–6 ions. The subsets with optimal predictive power have been found, and the ability to distinguish between methylene interrupted polyunsaturated fatty acids and other fatty acids have also been evaluated.

Keywords: Mass spectrometry, selected ion monitoring, fatty acid methyl esters, partial least squares regression, multiple linear regression.

1 Introduction

1.1 General

Fatty acid methyl esters (FAME) are the most common derivatives for gas chromatographic analyses of fatty acids. The application of FAME derivatives for the identification of unknown fatty acids by electron impact mass spectrometry (EI MS) has been limited because the spectra give incomplete information about the fatty acid structure, and in particular the number of double bonds in polyunsaturated fatty acids (polyenes) [1–3].

Determination of double bond positions can be carried out by derivatisation of the double bonds [4, 5] or by the derivatisation of the carboxyl group with nitrogen containing cyclic groups [4, 6]. These methods have certain drawbacks in common: 1) the preparation of the derivatives is more laborious and time consuming than the preparation of methyl esters. 2) The interpretation of the spectra often depends on the presence or absence of ions of rather low abundance. Preconcentration of the analyte may therefore be necessary to achieve spectra of sufficient quality. 3) The molecular weights of these derivatives are higher than for the corresponding FAME molecules. This usually leads to poorer chromatographic properties, and elution of long-chain fatty acids on polar GC columns may be a problem.

Interpretation of mass spectra is usually done by visual inspection and decisions are made on the presence or absence of certain critical ions. In the spectra of nitrogen containing fatty acid derivatives, double bond positions are for instance identified from diagnostic ions spaced by 12 amu instead of 14 amu. In EI mass spectra of methyl esters, polyunsaturated fatty acids (polyenes), diunsaturated fatty acids (dienes), monounsaturated fatty acids (monoenes) and saturated fatty acids, are distinguished by different base peaks. An alternative approach is to predict a fatty acid structure from smaller differences in relative abundances of all or selected ions in the spectra, by means of classification, regression, or other statistical analyses [7]. The statistical approach has proven beneficial in several cases. The identification of double bond position in monoenoic fatty acids is one example. It has been stated [3, 5, 8] that the double bond positions in these isomers can not be determined from EI spectra of FAME or other simple fatty acid esters, because the spectra have no ions indicating the positions. Nevertheless, it has been shown that the double bond positions can be determined from partial least squares regressions (PLSR) on EI spectra of fatty acid ethyl esters [9]. Similar approaches have been used for the determination of double bond positions in monounsaturated acetates [10, 11] and long-chain alcohols [10]. Another example is the identification of *trans* geometry in the unsaturated fatty acids [12, 13]. A further advantage of the statistical approach is that the reliability of the identification can be calculated from the error estimates of the regression or classification. Classical mass spectral interpretation provides no such information.

Correspondence: *Svein A. Mjøs*, Norwegian Institute of Fisheries and Aquaculture Research, Kjerreidviken 16, N-5141 Fyllingsdalen, Norway. Phone: +47-555-012-00, Fax: +47-555-012-99; e-mail: svein.mjos@fiskeriforskning.no

Traditional mass spectral interpretation of fatty acid derivatives requires a full-scan spectrum. However, monitoring of a few ions (selected ion monitoring, SIM) is advantageous both with respect to signal to noise ratio and scan speed. In EI spectra of FAME, ions of high abundance can be found in the lower mass range (m/z 50–110). The purpose of this work has been to show that fatty acid structure can be predicted from ions in this range and that there exist subsets of ions that are suitable for SIM analyses, with similar predictive power as full-scan spectra. The subsets of 4–6 ions with best predictive power have been sought.

If the number of double bonds in FAME is known, the rest of the fatty acid structure can often be predicted from retention indices and chromatographic elution patterns. The main focus has therefore been on prediction of the number of double bonds in methylene-interrupted (MI) polyenes, but the various subsets have also been evaluated for the ability to distinguish between MI polyenes, non methylene-interrupted (NMI) polyenes, dienes, monoenes and saturated fatty acids.

1.2 Electron impact mass spectra of FAME

The mass spectra of FAME and other simple esters have been thoroughly investigated and the identities of most of the abundant ions are known; several reviews have been published [1–3, 5]. Saturated unbranched FAME molecules are dominated by the ion $[CH_2C(OH)OCH_3]^{+\bullet}$ at m/z 74, which is formed via *McLafferty* rearrangement [14]. Other important ions are the series of ions with the formula $[(CH_2)_nCOOCH_3]^+$. The ion where n equals 2 (m/z 87) is usually the most abundant of the series, followed by ions where n is 6, 10 and 14 [5]. It has been shown that these ions can be formed by more than one mechanism [15]; a structure where the positive charge is located on protonated carbonyl oxygen after abstraction of the hydrogen from certain positions in the carbon chain, has been proposed [16]. Another important series in the saturated molecules are the ions with formula $[C_nH_{2n+1}]^+$ where m/z 43 and 57 are the most abundant ions [17].

In monoenes, a similar unsaturated series with formula $[C_nH_{2n-1}]^+$ dominates, and m/z 55 (n = 4) is usually the base peak in the spectra [17]. The series $[C_nH_{2n-3}]^+$ is abundant in MI dienes and m/z 67 (n = 5) is usually the base peak [17]. In MI polyenes with three or more double bonds the base peak is usually m/z 79, which belong to the series $[C_nH_{2n-5}]^+$.

As explained above, saturated FAME, monoenes, MI dienes, and MI polyenes can be distinguished by large differences among the most abundant ions in the spectra. The differences between the spectra of FAME with three

or more double bonds are much smaller, and m/z 79 is usually the base peak in all these spectra. A weak molecular ion may sometimes reveal the number of double bonds in the molecule. It has also been reported that the ion m/z 91, part of the series $[C_nH_{2n-7}]^+$, increase with increasing number of double bonds [8, 18].

The determination of the double bond position in polyunsaturated FAME can be done from diagnostic ions present in the EI spectra. The position of the first double bond counted from the methyl end of the carbon chain can be determined from a series of ions (ω-ions) with the molecular formula $[C_nH_{2n-4}]^{+\bullet}$ where m/z 108, 122, and 150 are indicative of n-3, n-4 and n-6 fatty acids respectively [1, 17, 19, 20]. A similar series of ions (α-ions) with the molecular formula $[C_nO_2H_{2n-6}]^{+\bullet}$ indicate the double bond positions counted from the carbonyl group, where Δ6 and Δ9 positions give ions of m/z 194 and 236 respectively [19].

2 Materials and methods

2.1 Overview

AOCS Ce 1b-89 [21] was used for preparation of fatty acid methyl esters from four commercial fish oils (cod liver, tuna, salmon, blue whiting) and a *Bligh* and *Dyer* [22] extract of blue mussels. The methyl esters were separated according to the degree of unsaturation by silver ion liquid chromatography (Ag-LC), and the fractions were analysed by GC-MS. After an initial screening, fractions containing fatty acids of interest (3–6 double bonds) were concentrated and rerun with several different GC-MS programs. Spectra were collected in a spectral base and calculations were performed on the spectral area of 50–110 amu. Further details are given below.

2.2 Silver ion chromatography (Ag-LC)

Ag-LC was performed on a *Chromspher* 5 Lipids column, 250 × 4.6 mm (*Varian*, Middelburg, The Netherlands) on a system equipped with evaporative light scattering detection and a fraction collector. Details are described elsewhere [13]. Fractions were collected every 2 min and the following gradient program was applied: Solvent A: hexane, Solv. B: acetone, Solv. C: 10% acetonitrile 90% acetone. 0–5 min: 100% Solv. A, 5–9 min: gradient 100% A to 80%A/20%B, 9–10 min gradient 80%A/20%B to 100%B, 10–30 min: gradient 100%B-100%C, 30–46 min: 100%C. Solvent flow was 1.5 mL/min. The split between the fraction collector and the detector was approximately 75/25. Saturated FAME and monoenes eluted in one peak. FAME with 2–6 double bonds eluted in different baseline separated peaks according to the number of double bonds.

2.3 Gas chromatography-mass spectrometry

The fractions were analysed on a HP-5890 gas chromatograph equipped with HP-5972 mass spectrometric detector (*Agilent*, Palo-Alto, CA, USA) and a BPX-70 capillary column, $L = 60$ m, *I.D.* $= 0.25$ mm, $d_f = 0.25$ μm (*SGE*, Ringwood, Australia). Two different temperature programs were applied to reduce the risk of undetected peak overlaps. To avoid influence on the mass spectra, column flow and MS transfer line temperature were equal for the two programs. Program 1: Injection (splitless) at 60 °C, isothermal for 4 min, thereafter 30 °C/min-190 °C and 4 °C/min-260 °C, isothermal (260 °C) for 5 min. The injector pressure was 414 kPa at injection; after 2 min the pressure was reduced to 138 kPa (73 kPa/min). Thereafter, the pressure was increased with column temperature (constant flow mode) to give an estimated carrier gas velocity of 20 cm/s. Helium (99.996%) were used as carrier gas. Injector temperature was 250 °C and MS transfer line temperature was 260 °C. The injected amounts were 0.5 or 1.5 μL, depending on the concentration of the compounds of interest. Program 2: Identical to program 1, with the following exceptions: the second temperature gradient (190–260 °C) was 2 °C/min, and there were no isothermal period after the second gradient. The two different GC-programs were combined with two MS-programs. 'Full scan' was applied for identification of the compounds from the molecular ion and the α- and ω-ions, m/z 50–382 were recorded, and 2.1 spectra were collected every second. 'Local scan' were used to get better signals in the region used for the calculations, m/z 50–110 were recorded, and 7.3 spectra were collected every second.

2.4 Identification

Because the fatty acids elutes according to the degree of unsaturation, the number of double bonds in a FAME can be estimated directly from Ag-LC retention times and from compounds with known structure appearing in the same Ag-LC fraction. In addition, the α- and ω-ions, the equivalent chain lengths (ECL), and fractional chain lengths (FCL) were used for identification. ECL values are retention indices calculated relative to the number of carbons in the saturated unbranched fatty acids [23, 24]. The FCL values are the difference between the ECL value of the actual FAME and the ECL value of the saturated unbranched FAME with the same number of carbons. All compounds applied in the calculations were identified by the presence of both the α- and ω-ions, except the *n*-1 fatty acids, which can not be identified from the ω-ion (see further details in results and discussion section).

A reference mixture, GLC-461 (*Nu-Chek Prep*. Elysian, MN, USA), spiked with additional saturated FAME, C19:0, C21:0, C25:0, C26:0, C27:0, and C28:0 (*Sigma-Aldrich*, Steinheim, Germany) was applied for calculations of ECL values. Local second order regressions were applied to determine the relationship between ECL values and retention times. Calculations were performed in *Matlab* 6.5 (*MathWorks*, Natick, MA, USA). A total of 30 fatty acids with 3–6 MI double bonds were identified, details are given in Tab. 1. For comparison of MI polyene spectra with other FAME spectra, the spectra in the reference mixture were included in the dataset together with two spectra tentatively identified as NMI polyenes. For the compounds with more than one spectrum in the original library, the average spectrum of the compound was calculated and the spectrum with best correlation to the average was applied further.

2.5 Calculations and error estimates

In the descriptions below, **X** is the matrix containing all spectra, and **y** is the vector containing the corresponding number of double bonds. Subsets of the spectra with 3–6 ions, $\mathbf{X}_s$, were tested for the ability to predict the number of double bonds, $\mathbf{y}_t$, in the fatty acids. Multiple linear regression (MLR) was used to find the vector of regression coefficients, **b**, from the spectral data, $\mathbf{X}_s$, and the number of double bonds in the molecule, $\mathbf{y}_t$:

$$\mathbf{b} = (\mathbf{X}_S^T \mathbf{X}_S)^{-1} \mathbf{X}_S^T \mathbf{y}_t^T \tag{1}$$

The predicted number of double bonds, $\mathbf{y}_p$, was then calculated from $\mathbf{X}_s$ and **b**:

$$\mathbf{y}_p = \mathbf{X}_s \times \mathbf{b} \tag{2}$$

Standard error of calibration (SEC) is the standard deviation of the difference between predicted values, $\mathbf{y}_p$, and true values, $\mathbf{y}_t$:

$$\mathrm{SEC} = \sqrt{\frac{1}{N-1} \sum_{n=1}^{N} (y_{p,n} - y_{t,n} - \delta)^2},$$
$$\delta = \frac{1}{N} \sum_{n=1}^{N} (y_{p,n} - y_{t,n}) \tag{3}$$

N is the number of cases.

Because of the risk of overfitting, error estimates calculated directly on the calibration set may be misleading. In full cross validation, each sample (fatty acid) is left out of the data set and the model is calibrated on the remaining samples. Then the y-value for the excluded sample is predicted from the model. The process is repeated with every sample in the calibration set. Standard error of prediction (SEP) is computed like SEC, but on cross validated $\mathbf{y}_p$. To resemble SIM spectra, all subsets (sub-spectra) were normalised to constant sum prior to the regressions.

Tab. 1. Polyunsaturated fatty acids applied in the calculations. ECL and FCL values are given together with the amount injected (estimated from the peak area compared to peak areas in the reference mixture).

Fatty acid	α-ion	ω-ion	ECL	FCL	ng inj.	Source
C16:3 *n*-3	264	108	17.94	1.94	0.4	Cod liver oil
C18:3 *n*-3[†]	292	108	19.99	1.99	19.6	Blue mussels
C20:3 *n*-3[†]	320	108	22.00	2.00	4.4	Reference compound[‡]
C22:3 *n*-3	348	108	24.00	2.00	7.2	Reference compound[‡]
C16:3 *n*-4	264	122	17.85	1.85	3.6	Cod liver oil
C18:3 *n*-4	292	122	19.86	1.86	1.8	Cod liver oil
C18:3 *n*-6[†]	292	150	19.65	1.65	25.0	Reference mixture
C20:3 *n*-6[†]	320	150	21.63	1.63	23.4	Reference mixture
C16:4 *n*-1	262	80[#]	18.56	2.56	7.3	Blue whiting
C18:4 *n*-1	290	80[#]	20.58	2.58	4.5	Salmon oil
C20:4 *n*-1	318	80[#]	22.60	2.60	1.2	Salmon oil
C16:4 *n*-3	262	108	18.25	2.25	6.8	Blue whiting
C18:4 *n*-3	290	108	20.47	2.47	46.7	Herring oil
C20:4 *n*-3	318	108	22.49	2.49	27.1	Salmon oil
C21:4 *n*-3	332	108	23.48	2.48	4.9	Blue whiting
C22:4 *n*-3	346	108	24.49	2.49	10.9	Blue whiting
C23:4 *n*-3	360	108	25.49	2.49	1.0	Blue whiting
C24:4 *n*-3	374	108	26.50	2.50	4.4	Blue whiting
C26:4 *n*-3	402	108	28.51	2.51	3.8	Blue whiting
C18:4 *n*-4	290	122	20.22	2.22	1.2	Blue whiting
C20:4 *n*-6[†]	318	150	21.99	1.99	49.4	Blue mussels
C21:4 *n*-6	332	150	23.10	2.10	0.5	Blue mussels
C22:4 *n*-6[†]	346	150	24.07	2.07	20.4	Reference mixture
C18:5 *n*-1	288	80[#]	20.94	2.94	1.7	Blue whiting
C20:5 *n*-3[†]	316	108	22.83	2.83	13.8	Blue whiting
C21:5 *n*-3	330	108	23.94	2.94	4.4	Blue whiting
C22:5 *n*-3[†]	344	108	24.94	2.94	21.2	Reference mixture
C24:5 *n*-3	372	108	26.97	2.97	3.7	Cod liver oil
C22:5 *n*-6	344	150	24.33	2.33	2.5	Cod liver oil
C22:6 *n*-3[†]	342	108	25.22	3.22	19.3	Reference mixture

[†] These compounds are present in the reference mixture.
[‡] Spectrum from pure (>99%) reference compound (*Nu-Chek Prep*).
[#] The ω-ion is not suitable for identification of the *n*-1 position. m/z 66 can be an alternative indicator. See text for details.

The optimal subsets of ions were found by testing every combination of 3–6 ions between m/z 50 and 110 for the ability to predict the number of double bonds in the fatty acids. The number of combinations, n_c, to test is given by:

$$n_c = \frac{a!}{b!(a-b)!} \qquad (4)$$

where a is the number of ions in the total dataset (m/z 50–110) and b is the number of ions in the subset. The subsets were tested by an iterative procedure programmed in *Matlab* 6.5. For computational reasons, SEC was used as initial error estimates. SEP for the ten solutions with lowest SEC was then calculated by full cross validation. In addition to searching for optimal combinations of all

ions, searches were also performed on two datasets where only ions with abundance of 10 and 20% of the base peak were included.

The predicted numbers of double bonds are rounded to the nearest integer. Assuming normally distributed residuals and negligible bias, the expected number of prediction errors, ENPE, can be estimated directly from SEP. ENPE equals the area percent under the normal distribution curve that is outside the 0.5 limits.

MLR can not be used in situations were there are more variables than compounds, or on datasets with large degree of correlation between the X-variables. Partial

least squares regression (PLSR) was therefore used for predictions based on all ions. In PLSR, the original X-data, are projected onto latent variables on which the regressions are performed. Further details about PLSR are given elsewhere [25–27]. Principal component analysis (PCA) was used to get a general overview over the structure in the data. PLSR with full cross-validation and PCA were performed in Unscrambler 6.0 (*CAMO*, Oslo, Norway). Spectra normalized to the base peak were used with PLSR.

3 Results and discussion

3.1 Description of the structure in the dataset

The compounds used in the MI polyene dataset are listed in Tab. 1, together with the source of the compound. Most of the compounds are well known constituents of marine lipids and were found in varying amounts in all samples. However, C21:4 *n*-3, C23:4 *n*-3, C24:4 *n*-3, C26:4 *n*-3 and C18:5 *n*-1 were only found in the blue whiting oil.

The average spectrum of the MI polyenes is shown in Fig. 1. The most abundant ion, m/z 79, was the base peak in all spectra, except in C20:3 *n*-6 and C18:3 *n*-4, where m/z 67 was the base peak. Other dominating ions are m/z 55, m/z 80, and m/z 93. Possible structures of these ions are given in Tab. 2 together with mean values and standard deviations. Several series of abundant ions spaced 14 amu apart can be seen in the average spectrum. The two most abundant series are m/z 53; 67; 81; 95; 109 and m/z 51; 65; 79; 93; 107. The different series are numbered in the Tab. The suggested structures in Tab. 2 are tentative and based on literature data (see introduction) and structures expected from the presence of
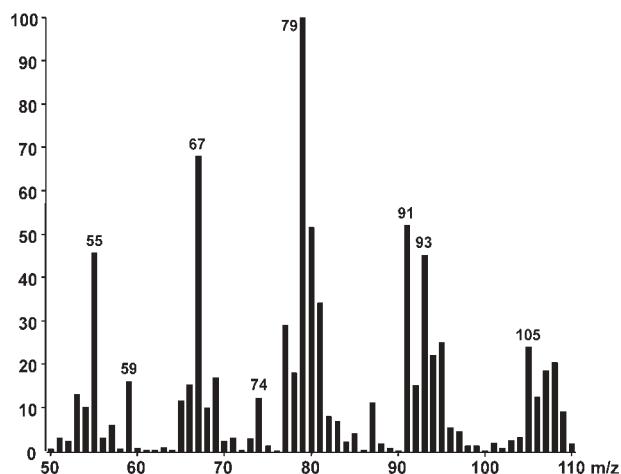
Tab. 2. Masses included in the subsets given in Tab. 3 and tentative identification of the ions. Mean and standard deviations are calculated on all MI polyenes in the dataset. Masses belonging to the same series are spaced by 14 amu (one methylene unit). Series 0 is a hypothetical ionized alkane; the series numbers increase with increasing loss of hydrogens.

Ion	Mean	Stdev.	Rel. Stdev. [%]	Series	Tentative identification
51	3.0	0.92	30.6	7	$[C_nH_{2n-5}]^+$ n = 4
52	2.4	0.66	28.2	6	$[C_nH_{2n-4}]^{+\bullet}$ n = 4
53	13.0	1.61	12.3	5	$[C_nH_{2n-3}]^+$ n = 4
54	10.1	3.97	39.4	4	
55	45.7	12.03	26.3	3	$[C_nH_{2n-1}]^+$ n = 4
57	6.0	3.11	52.0	1	$[C_nH_{2n+1}]^+$ n = 4
58	0.5	0.64	135.5	0	
64	0.3	0.56	183.1	8	
65	11.6	2.50	21.6	7	$[C_nH_{2n-5}]^+$ n = 5
66	15.3	5.14	33.6	6	$[C_nH_{2n-4}]^{+\bullet}$ n = 5 (Indicator for *n*-1)
67	68.1	14.35	21.1	5	$[C_nH_{2n-3}]^+$ n = 5
68	10.1	3.61	35.9	4	
69	17.0	6.79	40.0	3	$[C_nH_{2n-1}]^+$ n = 5
71	3.1	1.30	41.7	1	$[C_nH_{2n+1}]^+$ n = 5
74	12.3	3.58	29.2		*McLafferty* ion
77	29.0	5.88	20.3	9	Correlates to m/z 78. possibly $[C_6H_5]^+$
78	18.0	2.84	15.8	8	$[C_6H_6]^{+\bullet}$
79	99.9	0.54	0.5	7	$[C_nH_{2n-5}]^+$ n = 6 (Usually base peak)
80	51.5	13.45	26.1	6	$[C_nH_{2n-4}]^{+\bullet}$ n = 6
81	34.2	10.84	31.7	5	$[C_nH_{2n-3}]^+$ n = 6
83	6.8	3.03	44.5	3	$[C_nH_{2n-1}]^+$ n = 6
84	2.2	1.24	56.9	2	
91	52.1	15.98	30.7	9	$[C_nH_{2n-7}]^+$ n = 7
93	45.2	6.02	13.3	7	$[C_nH_{2n-5}]^+$ n = 7
94	22.0	6.58	29.9	6	$[C_nH_{2n-4}]^{+\bullet}$ n = 7
95	25.0	13.84	55.4	5	$[C_nH_{2n-3}]^+$ n = 7
96	5.3	3.42	64.1	4	
97	4.5	1.93	43.1	3	$[C_nH_{2n-1}]^+$ n = 7
99	1.2	0.51	41.4	1	$[C_nH_{2n+1}]^+$ n = 7
103	2.4	0.98	40.4		
104	3.2	2.12	66.0		
105	24.1	9.22	38.3	9	
106	12.5	7.59	60.8	8	
107	18.6	3.52	18.9	7	$[C_nH_{2n-5}]^+$ n = 8
108	20.4	12.75	62.5	6	$[C_nH_{2n-4}]^+$ n = 8 (Indicator for *n*-3)

Fig. 1. Average spectrum of the 30 methylene-interrupted polyenes in the dataset. All spectra were normalized to the base peak (100%).

related ions. It is likely that several of the abundant masses are caused by ions with the same mass but different molecular composition.

The ω-ions for the *n*-1 and *n*-3 series are m/z 80 and m/z 108 respectively. m/z 80 had a large abundance in all spectra and is not a suitable indicator for the *n*-1 class.

However, m/z 66 may be an alternative indicator. The abundance of m/z 66 was above 24% relative to the base peak for all *n*-1 fatty acids, while the abundance was below 17% for all other fatty acids. A possible structure for m/z 66 is $[C_5H_6]^{+\bullet}$, which can be formed by loss of a methylene unit from the ω-ion. All α-ions, and the ω-ions for *n*-4 and *n*-6, m/z 122 and 150 are outside the recorded mass range.

Principal component analysis (PCA) was used to get an overview over the data. Fig. 2A shows the 2 first principal components (PC) after PCA of the spectra of the MI polyenes, calculated together with spectra of saturated FAME, monoenes, dienes, and 2 compounds tentatively identified as NMI C20:3 and C22:3. The plot shows clear separation between saturated FAME, monoenes, dienes and polyenes. The 2 NMI polyenes are also separated from the MI polyenes. The NMI polyenes were found in the blue mussels sample and appeared in the same

HPLC-fraction as MI trienes. Weak molecular ions were present at m/z 320 and 348, and m/z 67 was the base peak in both spectra. ECL-values were 21.44 and 23.53 (GC program 2). The 2 compounds are probably 5,11,14-C20:3 and 7,13,16-C22:3, which have previously been identified in mussels [28].

Fig. 2B shows the score plot of the 2 first PCs when the PCA was performed only on the MI polyene dataset. The *n*-3, *n*-6 and *n*-4 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 between *n*-1 and *n*-3 was better when these two groups were calculated alone, also when the ω-ion (m/z 108) for the *n*-3 group was deleted from the dataset and the variables renormalised (plots not shown).

Within each group, there are sub-groups with the same number of double bonds. These are connected with lines. This means that there is some information about the num-
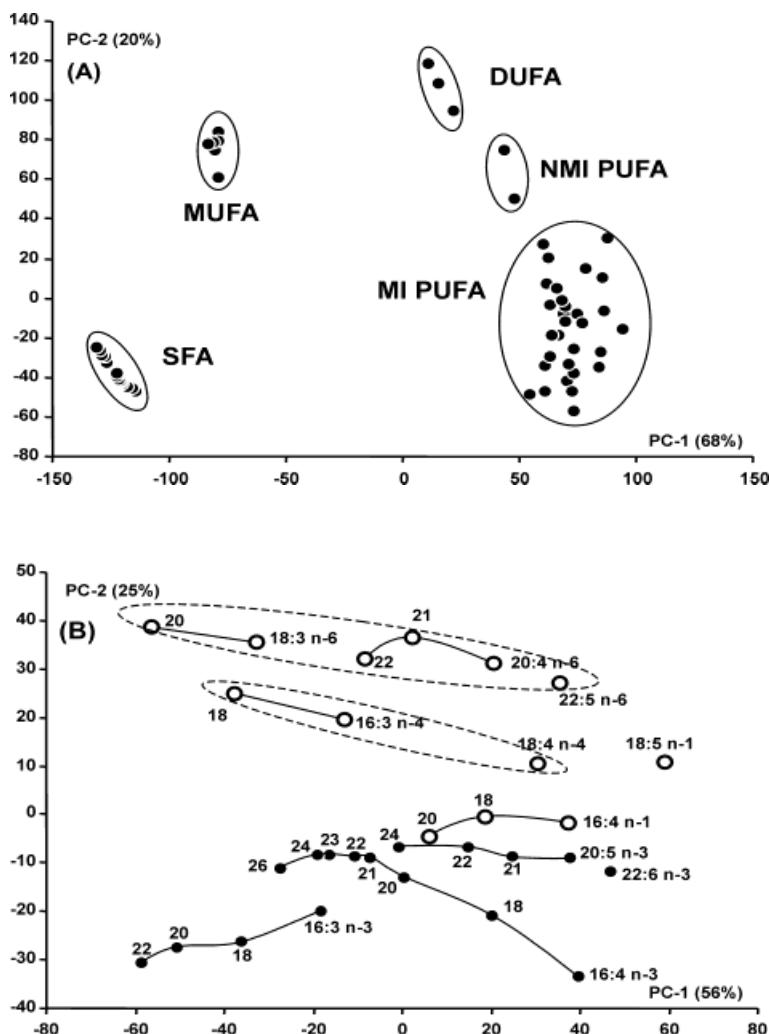

Fig. 2. Score plot of the 2 first principal components after PCA of spectra with all ions in the range m/z 50–110 used as variables. **A.** All spectra of the methylene-interrupted polyunsaturated fatty acids in the dataset (MI PUFA), together with saturated fatty acids (SFA), monounsaturated fatty acids (MUFA), and diunsaturated fatty acids (DUFA) from the reference mixture GLC-461. Included are also 2 fatty acids tentatively identified as non-methylene interrupted fatty acids (NMI PUFA). **B.** All spectra in the MI PUFA dataset. *N*-3 fatty acids are shown with closed circles, *n*-1, *n*-4 and *n*-6 fatty acids are shown with open circles. Classes of fatty acids with the same number of double bonds and the same *n*-x positions are connected with lines. The number of carbons is shown within each class.

ber of double bonds in the spectra, although it is not the dominating source of variation in the dataset. The separation along PC1 (56% of the variation) seems to be related to the position of the first double bond when counted from the carbonyl group (Δ-position). The fatty acids to the right in the plot are all Δ4 (C22:5 *n*-6, C22:6 *n*-3, C16:4 *n*-3) or Δ5 (C18:4 *n*-4, C18:5 *n*-1, C20:5 *n*-3, C20:4 *n*-6) polyenes.

3.2 PLS regression on all ions

PLSR was applied to build a model for the prediction of the number of double bonds based on all ions in the mass range 50–110 amu. The optimal number of PLS-components in the model was determined by full cross-validation. Predicted *vs*. true values are shown in Fig. 3A. SEP of 0.141 and bias of 0.01 gave an ENPE of 0.05%. All the predicted values were within the range of −0.5 to +0.5 of the true values; thus, there were no incorrect predictions. Although a correlation between the spectra and number of double bonds was not easily seen in Fig. 2B, the PLSR
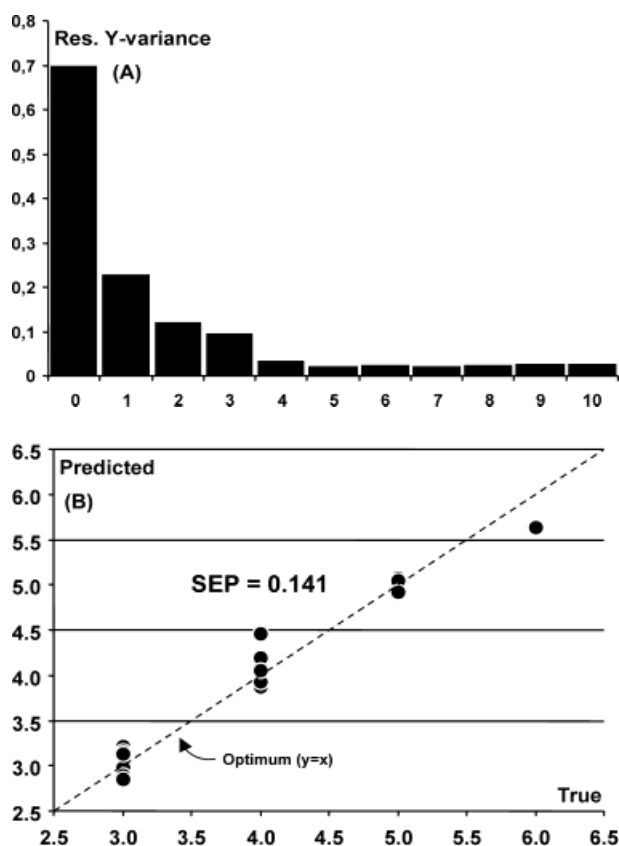

Fig. 3. PLS regression results for the prediction of the number of double bonds in MI polyenes from the EI mass spectra (m/z 50–110). **A.** Predicted *versus* true number of double bonds for the regression with 5 PLS components. **B.** Residual y-variance (cross validated) for models with 0–10 PLS components.

results show 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 MLR-model without loss of accuracy. The residual y-variance for PLS-models with different number of PLS-components is shown in Fig. 3B. The residuals falls rapidly from 0 to 4 components, there is a minimum at 5 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 3–6 ions.

3.3 Selection of ions for SIM

The results for the search for optimal ion combinations are shown in Tab. 3. For each search, the ten results with lowest SEC were cross validated. Only the five combinations with lowest SEP are given in Tab. 3. All searches showed positive correlation between SEP and SEC. With one exception, the model with lowest SEP was found among the models with the two lowest SEC values. This indicates that the models with lowest SEP will generally be among the ten models with lowest SEC. For the search for the five best combinations of all ions, the solution with best SEP had only the fifth best SEC. This search was therefore repeated and the 50 solutions with lowest SEC were cross validated. The five best solutions remained the same.

The bias in the models was too low to have any significance on ENPE. The estimation of ENPE assumes normally distributed residuals. The residual distribution was not checked for all models, but the sums of estimated and actual errors for all cross validated models were compared. The actual number of errors was 91% of the estimate. Regression of actual (y-variable) on predicted (x-variable) number of errors gave a slope of 0.87 and r^2 of 0.81.

Besides being able to predict the number of double bonds, an optimal subset of ions should also have other features. A strong signal is in most cases favourable. Expected signal strength was calculated for all combinations by summing the selected ions in the average spectrum (Fig. 1). The signal strength is given as percent relative to the signal strength of the n most abundant ions in the spectrum. Ideally, the selected subset should also be able to distinguish MI polyenes from other classes for which the MLR models are not valid. The ability to separate MI polyenes from saturated FAME, monoenes, dienes and NMI polyenes was tested by PCA, and visual inspection on score plots of the 2 first principal components, on the same dataset as applied in Fig. 2A. In cases

Tab. 3. Results for prediction of the number of double bonds in MI polyenes: Standard error of calibration (SEC), standard error of prediction (SEP), expected number of prediction errors (ENPE), Prediction errors, estimated signal strength, and class overlaps (see note) for the five best models with (**A**) 3 ions, (**B**) 4 ions, (**C**) 5 ions, (**D**) 6 ions. 10% abund. and 20% abund. denote that the subsets are selected among ions with at least 10% or 20% abundance relative to the base peak in the average spectrum (Fig. 1).

	No	SEC	SEP	ENPE (%)	Pred. Errors	Sign. Strength	Selected ions	Class overlap
A. 3 ions								
All ions	1	0.230	0.254	4.90	1	4.9	51, 96, 103	(1), 4
	2	0.237	0.264	5.82	1	11.7	54, 104, 106	(1)
	3	0.227	0.267	6.11	2	48.5	51, 80, 91	1, (2)
	4	0.256	0.285	7.94	2	25.9	51, 80, 103	1, 2
	5	0.253	0.296	9.12	3	48.2	52, 80, 91	1, (2), 4
10% abund.	6	0.252	0.297	9.23	2	52.4	65, 80, 91	(4)
	7	0.262	0.307	10.34	2	53.0	53, 80, 91	(1),
	8	0.270	0.308	10.45	4	78.4	79, 91, 108	1, (2), 3, 4
	9	0.267	0.314	11.13	2	60.3	77, 80, 91	1, 4
	10	0.276	0.316	11.36	3	79.8	79, 80, 105	1, 2, 3, 4
20% abund.	11	0.270	0.308	10.45	4	78.4	79, 91, 108	1, (2), 3, 4
	12	0.267	0.314	11.13	2	60.3	77, 80, 91	1, (2), 4
	13	0.276	0.316	11.36	3	79.8	79, 80, 105	1, (2), 3, 4
	14	0.289	0.339	14.02	4	54.9	80, 93, 105	1, (2), 3
	15	0.301	0.341	14.26	4	92.5	79, 80, 91	1, 2, 3, 4
Highest 3	16	0.371	0.412	22.49	7	100.0	67, 79, 91	
B. 4 ions								
All ions	17	0.170	0.201	1.29	0	6.5	51, 83, 96, 103	(2), (3)
	18	0.186	0.207	1.57	0	26.6	51, 66, 80, 103	1, 2, 3, 4
	19	0.188	0.213	1.89	1	21.4	51, 81, 103, 107	
	20	0.185	0.215	2.00	0	5.6	51, 96, 97, 103	
	21	0.189	0.217	2.12	0	8.5	51, 74, 96, 103	(2), (3)
10% abund.	22	0.193	0.232	3.11	2	68.3	68, 79, 80, 105	
	23	0.191	0.237	3.47	1	73.9	67, 77, 80, 91	
	24	0.198	0.237	3.49	1	81.2	69, 79, 80, 91	1
	25	0.210	0.246	4.21	3	68.0	79, 91, 106, 108	1, (2)
	26	0.205	0.249	4.46	3	61.4	55, 79, 80, 91	(1)
20% abund.	27	0.191	0.237	3.49	1	73.9	67, 77, 80, 91	
	28	0.205	0.249	4.46	3	91.8	55, 79, 80, 91	(1)
	29	0.204	0.252	4.72	2	77.2	79, 80, 81, 105	
	30	0.218	0.272	6.60	2	61.4	77, 80, 81, 91	
	31	0.233	0.276	7.00	2	82.4	79, 80, 91, 108	(3), (4)
Highest 4	32	0.308	0.366	17.19	5	100.0	67, 79, 80, 91	
C. 5 ions								
All ions	33	0.143	0.166	0.26	0	10.9	51, 54, 58, 103, 107	
	34	0.133	0.169	0.31	1	82.5	55, 79, 80, 91, 106	(1)
	35	0.136	0.170	0.33	0	72.0	65, 67, 80, 91, 93	(3)
	36	0.146	0.172	0.36	0	5.9	51, 83, 96, 99, 103	(2), (3), (4)
	37	0.148	0.173	0.39	0	10.8	51, 54, 64, 103, 107	2
10% abund.	38	0.133	0.169	0.31	1	82.5	55, 79, 80, 91, 106	(1)
	39	0.136	0.170	0.33	0	72.0	65, 67, 80, 91, 93	(3)
	40	0.143	0.180	0.55	0	84.2	55, 78, 79, 80, 91	
	41	0.143	0.183	0.63	0	63.4	65, 67, 78, 80, 91	
	42	0.153	0.189	0.82	1	73.4	69, 79, 80, 91, 106	(2)

Tab. 3. Continued.

	No	SEC	SEP	ENPE (%)	Pred. Errors	Sign. Strength	Selected ions	Class overlap
20% abund.	43	0.151	0.191	0.88	0	94.7	67, 77, 79, 80, 91	
	44	0.153	0.195	1.03	0	77.7	55, 67, 77, 80, 91	
	45	0.160	0.201	1.29	0	70.2	67, 77, 80, 91, 94	
	46	0.164	0.202	1.33	0	77.5	67, 77, 80, 91, 93	
	47	0.165	0.205	1.47	0	70.8	67, 77, 80, 91, 105	
Highest 5	48	0.210	0.273	6.70	3	100.0	55, 67, 79, 80, 91	
D. 6 ions								
All ions	49	0.101	0.132	0.02	0	84.7	55, 79, 80, 91, 93, 106	(1)
	50	0.110	0.135	0.02	0	50.5	71, 74, 79, 91, 104, 106	(1)
	51	0.116	0.148	0.07	0	63.6	65, 67, 80, 84, 91, 93	
	52	0.112	0.150	0.09	0	77.2	55, 78, 79, 80, 91, 106	
	53	0.115	0.150	0.09	0	62.2	57, 67, 77, 80, 91, 107	
10% abund.	54	0.101	0.132	0.02	0	84.7	55, 79, 80, 91, 93, 106	(1)
	55	0.112	0.150	0.09	0	77.2	55, 78, 79, 80, 91, 106	
	56	0.115	0.150	0.09	0	84.6	55, 74, 79, 80, 91, 93	
	57	0.117	0.155	0.13	0	68.2	55, 65, 67, 78, 80, 91	
	58	0.117	0.156	0.14	1	76.4	55, 66, 79, 80, 91, 106	
20% abund.	59	0.131	0.172	0.36	0	92.3	67, 77, 79, 80, 81, 91	
	60	0.136	0.176	0.45	0	77.4	55, 67, 77, 80, 81, 91	
	61	0.137	0.179	0.52	0	89.8	67, 77, 79, 80, 91, 95	
	62	0.140	0.190	0.85	0	87.3	55, 67, 77, 79, 80, 91	
	63	0.142	0.180	0.22	0	74.9	55, 79, 80, 91, 93, 94	(1)
Highest 6	64	0.165	0.216	2.06	0	100.0	55, 67, 79, 80, 91, 93	

Class overlaps: 1) Overlap between MI polyenes and NMI polyenes, 2) Overlap between MI polyenes and dienes, 3) Overlap between MI polyenes and monoenes, 4) Overlap between MI polyenes, and saturated FAME. Numbers in parenthesis indicates separation, but no clear grouping.

where the separation between the classes is poor, an alternative approach may be to identify the classes by means of retention time shifts [29] or by analysing the sample with an additional MS SIM program optimised for class separation.

All models based on 3 ions had SEP above 0.25 and ENPE of 5% or more. The class separation was also poor for these models. The only combination that could successfully separate MI polyenes from all other classes was the subset of the three most abundant ions, which had ENPE above 30%. The accuracy of the predictions improves as the number of ions in the subsets increase. Several of the subsets of 4 ions could successfully predict the number of double bonds for all compounds in the dataset. However, ENPE was still above 1%, and none of the subsets based on ions with average abundance above 10 or 20% was free of errors. ENPE is below 1% for most of the models based on 5 or 6 ions. Several of these models also have good signal strength and are able to distinguish MI polyenes from the other classes.

It should be emphasised that the error estimates given in Tab. 3 estimates the risk of model failure based on the spectra used in the regressions. Because the spectra in the applied dataset are acquired from samples that were fractionated and concentrated, they are assumed to be low in noise and free from interference from other compounds. In a practical situation there is always a risk of applying the models on spectra that are not suitable, either because they are influenced by noise, or because they are incorrectly classified as a MI polyenes. The risk of applying the model on an unsuitable spectrum is for obvious reasons not easily estimated. However, in many practical situations it can be expected to be at least 1%. Because of this risk, there may be little to gain by selecting subsets with ENPE below 1%. In the following discussion any model with ENPE below 1% is assumed to have sufficient accuracy.

Several of the models with 5 and 6 double bonds seem suitable from a practical point of view. ENPE is below, 1%, the signal strength is good, and the subset can be

used for classification purposes. A few of these subsets will be discussed in further detail. The 5-ion subset with lowest SEP is subset number 33. However, this subset includes an ion with average abundance below 1% (m/z 58). In addition to giving low signal strength, ions with this low abundance are easily influenced by noise; the model can therefore be expected to work only with spectra of very high quality. The second best model is acquired with subset 34; this subset is identical to the best subset of ions of average abundance above 10% (subset 38). ENPE is only 0.31, but the model failed to predict the number of double bonds in C16:4 *n*-3 (the predicted value was 4.58). The separation between MI polyenes and NMI polyenes was poor, and the classification between saturated FAME and monoenes was also incomplete. Subset 43, with the lowest SEP of the subsets based on ions with average abundance above 20%, is the only subset with signal strength above 90%. This subset gave excellent discrimination between MI polyenes and other classes, especially between MI polyenes and NMI polyenes, but saturated FAME overlapped with monoenes and polyenes.

Of the subsets with 6 ions, subset 49 gave lowest SEP. This is identical to the best subset of ions with average abundance above 10% (subset 54). The only difference between these subsets and subset 34/38 is the addition of m/z 93. The addition of this ion improved the regression properties, but the classification properties were not improved. Subset 50 and 56 are the only subsets that include the *McLafferty* ion, m/z 74. This is the base peak in saturated FAME, and both these subsets give excellent separation between saturated FAME, monoenes, dienes and polyenes. However, subset 50 gave poor separation between MI and NMI polyenes. This subset also contains 2 ions of low abundance, and can therefore not be recommended for most practical purposes. Subset 59 has the best signal strength and is identical to subset 43 with the addition of m/z 81. Similarly to subset 43, there is excellent discrimination between MI polyenes and the other classes, but there is overlap between saturated FAME and monoenes/dienes. Subset 61 differs from 59 by the replacement of m/z 81 with m/z 95, which belong to the same series (Series 5, Tab. 2). These two subsets have similar signal strengths and classification properties. Subset 62 is also similar to subset 59, but m/z 81 is replaced by m/z 55. Although m/z 74 is not included, this subset discriminates clearly between all classes, including saturated FAME.

Several of the subsets mentioned above can be suitable for the purpose of identification. For a quantitative purpose, subset 56 should be preferred. This subset includes the base peaks of saturated FAME (m/z 74),

monoenes (m/z 55) and polyenes (m/z 74). Dienes are also expected to give a good response because several of the selected ions are abundant in diene spectra. An alternative is to use one of the models without m/z 74 as a basis and include m/z 74. This will generally have no negative influence of the accuracy of the models. However, some care should be taken when working with real samples. m/z 74 is very weak in many polyene spectra, but dominating in spectra of saturated FAME. Chromatographic overlap between polyenes and even a very small peak of a saturated FAME will have strong influence on the abundance of m/z 74 in the polyene spectrum, and may therefore lead to incorrect predictions.

The MLR models applied in this study are based solely on all-*cis* MI polyenes. It has been shown that *trans* geometry in specific positions in MI polyenes has significant influence on the mass spectra [13]. The MLR models are therefore expected to fail if applied on certain *trans* isomers. Similarly, the models are expected to fail if applied on NMI polyenes. Even though the spectra of the 2 NMI polyenes could be distinguished from MI polyenes by PCA, the number of available compounds was too low to state this as a general rule. NMI polyenes with more than 3 double bonds may have a methylene-interrupted trienoic double bond system in addition to an isolated double bond. The spectra of such compounds can be expected to be more like the spectra of ordinary MI polyenes.

4 Conclusions

It has been shown that the number of double bonds in polyunsaturated fatty acid methyl esters with methylene-interrupted double bond systems can be determined from the mass region 50–110 amu in electron impact mass spectra. The number of double bonds could be predicted with high accuracy, both by partial least squares regression on all ions in this region, and by multiple linear regressions on selected subsets of 5–6 ions within this region. The subsets with optimal predictive power have been found, and the ability for class separation by these subsets has also been evaluated.

Acknowledgements

References

[1] *R. T. Holman, J. J. Rahm:* Analysis and characterization of polyunsaturated fatty acids. Prog. Chem. Fats Other Lipids **9** (1966) 13–90.

[2] *W. W. Christie:* Gas Chromatography and Lipids. The Oily Press, Ayr (Scotland) 1989, pp. 161–184.

[3] *G. Dobson, W. W. Christie:* Spectroscopy and spectrometry of lipids – Part 2. Mass spectrometry of fatty acid derivatives. Eur. J. Lipid Sci. Technol. **104** (2002) 36–43.

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

[5] *K. K. Sun, R. T. Holman:* Mass spectrometry of lipid molecules. J. Am. Oil Chem. Soc. **45** (1968) 810–817.

[6] *W. W. Christie:* Gas-Chromatography-mass spectrometry methods for structural analysis of fatty acids. Lipids **33** (1998) 343–353.

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

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

[9] *F. Brakstad*: 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** (1993) 87–100.

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

[11] *B. S. Lanne, M. Appelgren, G. Bergström:* Determination of the double bond position in monounsaturated acetates from their mass spectra. Anal. Chem. **57** (1985) 1621–1625.

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

[13] *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.

[14] *F. W. McLafferty:* Molecular rearrangements. Anal. Chem. **31** (1959) 82–87.

[15] *N. Dihn-Nguyên, R. Ryhage, E. Stenhagen:* Mass spectrometric studies. VIII. A study of the fragmentation of normal long-chain methyl esters and hydrocarbons under electron impact with the aid of deuterium-substituted compounds. Arkiv För Kemi. **18** (1961) 393–399.

[16] *G. Spiteller, M. Spiteller-Friedman, R. Houriet:* Klärung massenspectrometrischer Zerfallsmechanismen durch Verwendung kalter Ionenquellen und von Elektronen niedriger Energie, 1. Mitt: Aliphatische Ester. Monatsh. Chem. **97** (1966) 121–128.

[17] *J. J. Myher, L. Marai, A. Kuksis:* Identification of fatty acids by GC-MS using polar siloxane liquid phases. Anal. Biochem. **62** (1974) 188–203.

[18] *G. Odham, E. Stenhagen:* Fatty acids. In: Biochemical applications of mass spectrometry. Ed. *G. R. Wallace*, Wiley, NY (USA) 1972, pp. 211–228.

[19] *H. Brauner, Budziewicz, W. Boland:* Studies in chemical ionization mass spectrometry. Org. Mass Spectrom. **17** (1982) 161–164.

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

[21] *D. Firestone:* Official Methods and Recommended Practices of the American Oil Chemists' Society, 5th edn. American Oil Chemists' Society, Champaign, IL (USA) 1998.

[22] *E. G. Bligh, W. J. Dyer:* A rapid method of total lipid extraction and purification. Can. J. Biochem. Physiol. **37** (1959) 911–917.

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

[24] *F. P. Woodford, C. M. van Gent:* Gas liquid chromatography of fatty acid methyl esters: the "carbon number" as a parameter for comparison of columns. J. Lipid. Res. **1** (1960) 188–191.

[25] *E. N. Malinowski:* Factor analysis in chemistry, 3rd edition. Wiley, NY (USA) 2002.

[26] *P. Geladi, B. R. Kowalski:* Partial least-squares regression: A tutorial. Anal. Chim. Acta. **185** (1986) 1–17.

[27] *L. Lorber, Wangen, B. R. Kowalski:* A theoretical foundation for the PLS algorithm. J. Chemometrics. **1** (1987) 19–31.

[28] *J. L. Garrido, I. Medina:* Identification of minor fatty acids in mussels (Mytilus galloprovincialis) by GC-MS of their 2-alkenyl-4,4-dimethyloxazoline derivatives. Anal. Chim. Acta. **465** (2002) 409–416.

[29] *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.