https://doi.org/10.1016/j.foodchem.2020.127601

Quantitative Analysis of Blended Oils by Matrix-assisted Laser Desorption/Ionization Mass Spectrometry and Partial Least Squares Regression

Suying Li,[a,b] Tsz-Tsun Ng[a.b] and Zhong-Ping Yao[a,b,*]

[a]State Key Laboratory of Chemical Biology and Drug Discovery, Food Safety and Technology Research Centre and Department of Applied Biology and Chemical Technology, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong Special Administrative Region, China

[b]State Key Laboratory of Chinese Medicine and Molecular Pharmacology (Incubation) and Shenzhen Key Laboratory of Food Biological Safety Control, Shenzhen Research Institute of Hong Kong Polytechnic University, Shenzhen 518057, China

*To whom correspondence should be addressed:

Zhong-Ping Yao
Department of Applied Biology and Chemical Technology
The Hong Kong Polytechnic University
Hung Hom, Kowloon
Hong Kong
Tel: (852) 34008792. Fax: (852)23649932.
E-mail: zhongping.yao@polyu.edu.hk

Abstract

Quantitative labeling of oil compositions has become a trend to ensure the quality and safety of blended oils in the market. However, methods for rapid and reliable quantitation of blended oils are still not available. In this study, matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) was used to profile triacylglycerols in blended oils, and partial least squares regression (PLS-R) was applied to establish quantitative models based on the acquired MALDI-MS spectra. We demonstrated that this new method allowed simultaneous quantitation of multiple compositions, and provided good quantitative results of binary, ternary and quaternary blended oils, enabling good limits of detection (e.g., detectability of 1.5% olive oil in sunflower seed oil). Compared to the conventional GC-FID method, this new method could allow direct analysis of blended oils, analysis of one blended oil sample within minutes, and accurate quantitation of low-abundant oil compositions and blended oils with similar fatty acids contents.

Keywords: Blended oils; quantitation; mass spectrometry; matrix-assisted laser desorption/ionization; partial least squares regression.

1. Introduction

Vegetable oils are commonly used in our daily life. The average consumption of vegetable oils per capita per year was reported to be 17.3 kg in India, 25.3 kg in China, and 25.8 kg in Europe in 2015-2017 (OECD/FAO, 2018). Intake of suitable fatty acids (FAs) was recommended by the World Health Organization (WHO) for a healthy diet, with the intake ratio of saturated FAs, monounsaturated FAs and polyunsaturated FAs around 1:1.5:1, and the intake ratio of linoleic acid (omega 6) and alpha linolenic acid (omega 3) between 5:1 and 10:1 (FAO, 2010). The FAs profiles of most pure vegetable oils on their own fail to meet the requirements individually. Therefore, blended oils have become more and more popular because of their higher nutritional values, enhanced flavors and advantages in physical and chemical properties (Choi, Lee, & Lee, 2014; Choudhary, Grover, & Kaur, 2015). Because of the enormously varied prices of different pure vegetable oils, mislabeling of the oil compositions for financial gain, particularly exaggerated labeling of those more expensive oils, e.g., olive oil and grapeseed oil, has been a problem frequently encountered in the market of blended oils (Roxborough, 2018; Tien & Hsu, 2017). Regulations about this have been made. For example, the percentages of all the contained pure oil compositions are required to be labeled for blended oil products in China and India (FSSAI, 2011; NHC & SAMR, 2018), and in the European Union, the percentage of olive oil is required to be indicated for olive oil-containing blended oil products (EC, 2012). Reliable methods for quantitative analysis of blended oils are required to meet the numerous analytical demands for quantitative labeling and quality control of blended oil products.

Gas chromatography (GC) is the conventional method for analyzing edible oils. After the conversion of triacylglycerols (TAGs), the predominant chemical composition of edible oils, to fatty acid methyl esters (FAMEs) by chemical derivatization, the FAs contents of edible oils could be determined using

GC coupled with flame ionization detector (FID) or mass spectrometry (MS). Pure edible oils could be identified by matching the obtained FAs contents with the Codex standards (Codex, 2015), and the concentrations of the individual oils in blended oils were determined using chemometric tools to process the obtained FAs contents (Monfreda, Gobbi, & Grippa, 2014; Xie, Liu, Yu, Song, & Hu, 2013). TAGs can also be directly analyzed using high-temperature gas chromatography (Park, Chang, & Lee, 2010; Ruiz-Samblas, Marini, Cuadros-Rodriguez, & Gonzalez-Casado, 2012). High-performance liquid chromatography (HPLC) have been more commonly used to profile TAGs contents of edible oils, and quantitation of olive oil in blended oils could be achieved based on the abundances of selected TAG peaks (Fasciotti & Pereira Netto, 2010) in chromatograms and chemometric analysis (de la Mata-Espinosa, Bosque-Sendra, Bro, & Cuadros-Rodríguez, 2011). Both the GC and HPLC approaches require sample pretreatment and column separation which could be laborious and time-consuming (Cozzolino & De Giulio, 2011). Direct analysis techniques, including those based on fluorescence (Li, Wang, Zhao, Ouyang, & Wu, 2015; Milanez, Nóbrega, Nascimento, Insausti, Band, & Pontes, 2017; Poulli, Mousdis, & Georgiou, 2006), UV-visible (Milanez, Nóbrega, Nascimento, Insausti, Band, & Pontes, 2017), Raman (El-Abassy, Donfack, & Materny, 2009) and Fourier transform infrared spectroscopy (FTIR) (Li, Wang, Zhao, Ouyang, & Wu, 2015; Rohman & Che Man, 2012), and electrospray ionization mass spectrometry (Alves, Sena, & Augusti, 2014), have been used for quantitative analysis of blended oils. However, most of the previous studies on oil quantitation focused on binary blended oils, with very few on blended oils with multiple compositions, which are very common in the market (Jović, Smolić, Primožič, & Hrenar, 2016; Rohman & Che Man, 2010, 2011). For blended oils with more compositions, the possible combinations of oil compositions become more complicated, resulting in a significantly increase in the number of samples for establishing the calibration relationship, more complex spectral data and consequently

more challenging quantitative analysis.

Matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) is a technique with advantages such as simple sample preparation, short analysis time, high tolerance to impurities and high-throughput capacity (Cozzolino & De Giulio, 2011; Ng, So, Zheng, & Yao, 2015). It has been shown that MALDI-MS could detect the TAGs contents of edible oils (Ayorinde, Elhilo, & Hlongwane, 1999; Lay, Liyanage, Durham, & Brooks, 2006; SCHILLER, SÜß, PETKOVI, & ARNOLD, 2002), which allows the classification of edible oils (Calvano, Palmisano, & Zambonin, 2005; Kuo, Kuei, Hsiao, Chung, Hsu, & Chen, 2019). We have previously established a simple protocol for direct analysis of edible oils (Ng, Li, Ng, So, Wong, Li, et al., 2018; Ng, So, Zheng, & Yao, 2015). Our previous study showed that the intensity ratio of selected ions could be used to determine the concentrations of olive oil in blends of olive oil and canola oil (Ng, So, Zheng, & Yao, 2015), but such an approach was not applicable for blended oils with multiple compositions. Partial least squares regression (PLS-R) is a commonly used multivariate regression method, which is powerful in dealing with complex and multicollinear data (Jović, Smolić, Primožič, & Hrenar, 2016; Monfreda, Gobbi, & Grippa, 2014). PLS-R extracts the most useful information (called PLS components) from original data, builds models using the PLS components to reduce data complexity and provides multiple predictors simultaneously based on the dimension-reduced models (Wold, Sjöström, & Eriksson, 2001). In this study, we aimed to develop a method for rapid quantitation of blended oils using MALDI-MS. We hypothesized that the MALDI-MS spectrum of a blend oil contained the quantitative information of its compositions and such information could be obtained using the PLS-R models. By investigating various blended oils, including binary, ternary and quaternary blends as well as commercial products, we demonstrated for the first time that

simultaneous quantitation of multiple compositions in blended oils could be achieved by using MALDI-MS and the developed method was simple, high-throughput and had high sensitivity in the measuring of oil compositions.

2. Materials and Methods

2.1 Chemicals

2, 5-Dihydroxybenzoic acid (DHB) and α-cyano-4-hydroxycinnanic acid (CHCA) were ~~was~~ purchased from Aldrich (St. Louis, MO, USA). HPLC grade acetone, HPLC grade methyl tert-butyl ether (MTBE), and trimethylsulfonium hydroxide (TMSH, 0.25 M solution in methanol) were purchased from Acros Organic (Waltham, MA, USA). HPLC grade acetonitrile (ACN) was purchased from Anaqua Chemical Supply (Houston, TX, USA). Polyethylene glycol (PEG) standards were purchased from Fluka (St. Louis, MO, USA) and sodium iodide (NaI) was purchased from Panreac Química (Barcelona, Spain). The mixture standard of 37 FAMEs was purchased from Supelco (St. Louis, MO, USA).

2.2 Oil samples

Vegetable oil products were collected from suppliers in Hong Kong and in mainland China, with the details as shown in Table S1. In this study, both commercial blended oil products as well as blended oil samples, which were prepared in laboratory by mixing pure edible oils in different weight ratios (Table S2), were analyzed. For each commercial blended oil product, the corresponding pure oil products of the same brands were also collected. Blended oil samples were divided into a training set and a testing set for establishing and validating PLS-R models, respectively. For each type of binary blended oils, there were 2 pure samples, and 4 blended samples with an increment of 20% in

concentrations (i.e., 20%:80%, 40%:60%, 60%:40% and 80%:20%) in the training set; the testing set had 5 blended samples that were not in the training set. For each type of ternary blended oils, there were 66 samples with an increment of 10% in concentrations in the training set, namely 3 pure samples, 9 × 3 binary blended samples, and 36 ternary blended samples (e.g., 0%:0%:100%, 0%:10%:90% and 10%:10%:80%); the testing set had 12 samples including pure and blended samples. For each type of quaternary blended oils, there were 286 samples with an increment of 10% in concentrations in the training set, namely 4 pure samples, 9×6 binary blended samples, 36×4 ternary blended samples, and 84 quaternary blended samples (e.g., 0%:0%:0%:100%, 0%:0%:10%:90%, 0%:10%:10%:80% and 10%:10%:10%:70%); the testing set had 13 samples including pure and blended samples. All the oil samples were sealed and stored in a dark and dry place before analysis.

2.3 MALDI-MS analysis

Sample preparation for MALDI-MS analysis was performed using a previously reported protocol (Ng, et al., 2018). Briefly, aliquots of 0.5 μL of 100 mg mL^{-1} DHB in acetone were loaded onto spots of the MADLI plate and air-dried. Each oil sample was directly applied as a thin layer on the DHB by using a medical cotton tip. PEG solution mixture (PEG600/PEG1000/PEG2000/NaI = 1/2/2/5 (v/v)) was mixed with 10 mg mL^{-1} CHCA solution (ACN/H_2O = 7/3 (v/v)) and then loaded onto the MALDI plate for calibration of the mass spectrometer.

An UltrafleXtreme MALDI-TOF/TOF mass spectrometer (Bruker, Billerica, MA, USA) equipped with a 355 nm smartbeam-II laser was operated in positive ion and reflectron mode for the MALDI-MS analysis. The ion source voltage 1, ion source voltage 2, lens voltage, reflector voltage 1 and reflector voltage 2 were set to 19 kV, 16.4 kV, 8 kV, 21 kV and 9.6 kV, respectively. The ion pulse

excitation was set to 80 ns. Mass spectra with a *m/z* range of 500–2000 Da were acquired automatically with the laser intensity varied in the limited range, and each shot included 1000 laser pulses. 8 single spectra with resolutions higher than 3000 in the TAGs range (typically at the region of *m/z* 850–920) were accumulated and saved as one spectrum for further analysis. Each sample was analyzed in eight replicates.

2.4 GC-FID analysis

Conversion of oil samples to FAMEs for GC-FID analysis was made according to ISO 12966-3 (ISO, 2014). The prepared FAMEs samples were analyzed by a 6890N gas chromatograph with an Agilent DB-23 column (60 m, 0.25 mm i.d., 0.25 μm film thickness) and a flame ionization detector (Agilent, Santa Clara, CA, USA). The injector temperature and detector temperature were 250 ºC and 280 ºC, respectively. The temperature of the oven was set at 50 ºC at the beginning, held for 1 min, and increased to 175 ºC at 25 ºC min^{-1}. Then the oven temperature was slowly increased from 175 ºC to 230 ºC at 3 ºC min^{-1} and held at 230 ºC for 10 mins. The gas flow of nitrogen and hydrogen was 4 mL min^{-1} and 40 mL min^{-1}, respectively. Each sample was analyzed in triplicate, and the chromatograms of the oil samples were compared with those of the mixture standard of 37 FAMEs to determine the fatty acid contents.

2.5 Statistical analysis

The MALDI-MS spectra and GC-FID chromatograms were normalized by making the total intensities of all the chosen peaks in one profile as 100% before statistical analysis. The normalized intensities of the chosen peaks and corresponding concentrations of oil compositions were input into a statistics software (Umetrics Simca 14.0, Andover, MA, USA) to establish the PLS-R models.

The protocol for establishing the PLS-R models (Wold, Sjöström, & Eriksson, 2001) is shown in Figure S1, with five major steps involved. In the 1st step, the number of PLS components (A) is a critical factor for the performance of PLS-R model, and in this study, Wold's R criterion (Li, Morris, & Martin, 2002) and RMSE criterion (Jović, Smolić, Primožič, & Hrenar, 2016) (see Section S1 of Supporting Information for the details) were applied to obtain optimal A and prevent over-fitting. In the 2nd step, the PLS-R model was evaluated by two parameters, i.e., R^2Y and Q^2, which described the fitting ability and predictive ability of the model, respectively. In the 3rd step, outliers with standardized residuals greater than +4 or -4 standard deviations (SDs) and "unimportant" variables with VIP (variable importance for the projection) values lower than 0.5 were excluded from the original data. A new fitting based on the reduced data set was carried out by repeating the above procedures. When a PLS-R model with good R^2Y and Q^2 values (both > 0.8 in this study) (Cho, Yang, Kim, Kim, Ko, Riu, et al., 2009; Veerasamy, Rajak, Jain, Sivadasan, Varghese, & Agrawal, 2011) but without outliers and unimportant variables was obtained, the model fitting was finished. To review the predictive ability of the established PLS-R model, the compositions of samples in both the training set and the testing set were quantified by the model. In the 4th step, the root mean square error of estimation (RMSEE), root mean square error of cross-validation (RMSEcv) and correlation coefficient (R^2) were used to describe the difference between the actual concentrations and the measured concentrations of samples in the training set. In the 5th step, the root mean square error of prediction (RMSEP) and correlation coefficient (R^2) were used to describe the difference between the actual concentrations and the measured concentrations of samples in the testing set. Grubbs test with detection level $\alpha = 0.05$ was carried out to detect outliers of the measured results of testing samples.

3. Results and discussion

3.1 MALDI-MS spectra of blended oils

Olive oil and sunflower seed oil blends are one of the most common blended oil products in the market due to their improved quality and high nutritional value. Figure 1 shows the MALDI-MS spectra of pure olive oil, pure sunflower seed oil and blends of olive oil and sunflower seed oil in different blending ratios, with the peaks assigned according to the literatures (Calvano, Palmisano, & Zambonin, 2005; Lay, Liyanage, Durham, & Brooks, 2006; Ng, et al., 2018). Obvious changes of peak abundances in the TAGs region were observed with the varied oil compositions. The highest intensity peaks in the spectrum of pure sunflower seed oil (Figure 1a) were *m/z* 901.7 (LLL, L: linoleic acid) and *m/z* 903.7 (OLL, O: oleic acid), while olive oil (Figure 1d) was abundant with *m/z* 881.8 (POO, P: palmitic acid) and *m/z* 907.8 (OOO). With increased concentration of olive oil in the oil blends, the relative abundances of *m/z* 881.8 and *m/z* 907.8 increased and the peaks at *m/z* 901.7 and *m/z* 903.7 became lower, indicating a correlation between the oil compositions of the blended oils and their MALDI-MS spectra.

3.2 Quantitative analysis of binary blended oils

3.2.1 Determination of optimal A

To quantify the oil compositions of olive oil and sunflower seed oil blends, PLS-R models were established based on the acquired MALDI-MS spectra. As described in Section 2.5 and Supporting Information, A is critical for the performance of PLS-R model. Wold's R criteria with thresholds at 0.90 (called R(0.90)) and 0.95 (called R(0.95)) and RMSE criterion were applied to determine the optimal A, and factors of the corresponding PLS-R models were summarized in Table S3. The optimal A of R(0.90), R(0.95) and RMSE increased in turn, leading to better fitting ability (decreased RMSEE)

and better predictive ability (decreased RMSEcv and RMSEP) of the PLS-R models. For PLS-R models of ternary blended oils (Table S4), the same optimal A was obtained for R(0.90) and R(0.95), which was smaller than that of RMSE. The RMSE model showed better fitting ability compared with the R(0.90) and R(0.95) models, while the RMSEP values of the RMSE model (0.0282, 0.0358 and 0.0298 for olive oil, perilla oil and sunflower seed oil, respectively) were similar to that of the R(0.90) and R(0.95) models, indicating that there were no differences in the predictive abilities among the models (Table S2). For blended oils containing four oil compositions (Table S5), the optimal A of the R(0.90) model, the R(0.95) model and the RMSE model significantly increased, leading to the improved fitting ability. Comparing the R(0.90) model with the R(0.95) model, better quantitative results (smaller RMSEP) of all the oil compositions were observed from the R(0.95) model which had a larger A; while for the R(0.95) model and the RMSE model, there was little improvement or even slight deterioration in the quantitative results of some compositions when A increased from 10 to 15, which might be related to over-fitting. In this study, R(0.90) tended to produce smallest A while RMSE usually generated models with the largest A among the three criteria. ~~Therefore~~ Overall, R(0.95) was more suitable for the determination of optimal A to prevent over-fitting and achieve a balance between fitting ability and predictive ability of the PLS-R models.

3.2.2 Validation of the PLS-R model

According to the R(0.95) criteria, the optimal **A** of the PLS-R model of olive oil and sunflower seed oil blends was determined as 2 (Table S3), and the olive oil concentrations of samples in the testing set were measured by the model to validate its predictive ability. For all the testing samples with olive oil concentrations from low to high levels, the relative errors of the measured results was close to 0 and the intra-day precision was in the range of 0.4–4.9% (Table S6), illustrating that the PLS-R could

be applied to quantify the oil compositions of olive oil and sunflower seed oil blends. Based on the established PLS-R model and the measured results of the blank sample (i.e., the pure sunflower seed oil), the limit of detection (LOD) of olive oil was calculated as 0.9% (see Section S2 of Supporting Information for the calculation) (Olivieri, Faber, Ferré, Boqué, Kalivas, & Mark, 2006; Ortiz, Sarabia, Herrero, Sánchez, Sanz, Rueda, et al., 2003). Two blended samples with 1.0% and 1.5% olive oil were prepared, and they were measured to contain 0.7% and 1.0% olive oil, respectively, by the model. Therefore, the PLS-R model of olive oil and sunflower seed oil blends could experimentally detect the presence of olive oil from at least down to 1.5% in sunflower seed oil. To investigate the reproducibility of the developed approach, the same testing samples were analyzed for 8 selected days (i.e., 1st, 2nd, 3rd, 5th, 7th, 10th, 14th and 20th) in 20 days and their olive oil concentrations were measured by the PLS-R model established on the first day (Table S7). The largest relative errors (–16.1%) and largest RSD (11.8%) belonged to the sample with low olive oil concentration, and for the samples with higher concentrations, the relative errors was between –7.0% and –0.1% with inter-day precision within 1.6-2.0% (Table S6), demonstrating the durable applicability of the established models.

The quantitative analysis of olive oil and sunflower seed oil blends has been previously investigated using fluorescence (Poulli, Mousdis, & Georgiou, 2006) and Raman (El-Abassy, Donfack, & Materny, 2009) spectroscopy coupled with PLS-R. For pure olive oil and sunflower seed oil, noticeable differences were observed from their fluorescence spectra, while the differentiation of these two oils in their Raman spectra mainly depended on the differences in peak intensities. For the fluorescence-based PLS-R model, the RMSEE and RMSEP were 1.7% and 3.4%, respectively, while the Raman-based PLS-R method showed RMSEE and RMSEcv as 2.56% and 3.59%, respectively. All these

parameters were larger than the corresponding parameters of the MALDI-MS-based PLS-R model (0.52%, 0.56% and 0.58% for RMSEE, RMSEcv and RMSEP , respectively, as shown in Table S3), indicating that the MALDI-MS-based method could provide more accurate results for quantitation of olive oil and sunflower seed oil blends.

3.3 Quantitative analysis of blended oils with multiple compositions

3.3.1 Quantitative analysis of ternary and quaternary blended oils

Ternary and quaternary blended oils are commonly available in the market, but such multiple compositions make quantitative analysis of their compositions much more complicated. Olive oil, perilla oil and rice bran oil are often added into blended oil products because of their special nutritional characteristics, and sunflower seed oil is a common composition of blended oil products. In this study, MALDI-MS integrated with PLS-R was employed to analyze ternary and quaternary blended oils, with olive oil, perilla oil and sunflower seed oil blends as well as olive oil, perilla oil, rice bran oil and sunflower seed oil blends as examples. Their PLS-R models were established following the protocol (Figure S1) and R(0.95) as described in Section 2.5.

The model of olive oil, perilla oil and sunflower seed oil blends showed the best description and prediction to olive oil with the lowest RMSEE, RMSEcv and RMSEP. The highest RMSEP was obtained at perilla oil, indicating that the predictive ability of the model for perilla oil was slightly worse than that for the other two compositions (Table S4). For olive oil, perilla oil, rice bran oil and sunflower seed oil blends, the PLS-R model showed good fitting ability ($\leqslant$0.0170) and powerful predictive ability ($\leqslant$0.0221) to olive oil and perilla oil, and the largest RMSEcv (0.0243) and RMSEP (0.0320) belonged to rice bran oil and sunflower seed oil, respectively (Table S5). As the number of

oil compositions increased, the number of training samples significantly increased. For the PLS-R model of the ternary blended oils, 66 samples in the training set formed a large triangle in the score plot (Figure 2a), with the three vertices of the triangle representing the three pure oils, the three edges formed by the binary blends of the three pure oils, and the interior of the triangle filled with the ternary blended oils. For the quaternary blended oils, the training set of the PLS-R model contained 286 oil samples, forming a tetrahedron in the 3D score plot with the faces for the ternary blended oils and the inside for the quaternary blended oils (Figure 2b).

To validate the predictive ability of PLS-R models for ternary and quaternary blended oils, the compositions of testing samples were quantified by the established models. As shown in Table 1, for the olive oil, perilla oil and sunflower seed oil blends, the measured results for high-abundant (>20%) compositions were close to their actual concentrations, with the relative errors and precision within –16.7-12.0% and 0.5-4.7%, respectively. For the quaternary blends of olive oil, perilla oil, rice bran oil and sunflower seed oil, the PLS-R model showed satisfactory predictive ability to all the high-abundant compositions with the relative errors and precision within –19.9-10.4% and 0.5-9.2%, respectively (Table S8). However, for the low-abundant compositions (~5%) of the ternary blended oils, the relative errors and precision of the measured results varied from –45.0% to 41.9% and 2.6% to 38.9%, respectively, indicating the poor predictive ability of the PLS-R model for the compositions at low levels. Meanwhile, there was the possibility of false positive results of compositions that were not present, which was observed the quantitative results of both the ternary and quaternary blended oils.

For ternary and quaternary blended oils, the multiple oil compositions caused complex spectral data

and very large numbers of training samples (66 and 286 for ternary and quaternary blends, respectively). Since the PLS-R models were established based on all the training samples, the maximum differences of the training samples (i.e., differences between the pure oils) were summarized and some minor differences that were critical to differentiate compositions at low levels might be ignored by the models, leading to the poor predictive ability of the PLS-R models for the low-abundant compositions. To overcome such drawbacks, for the first time, a two-step zoom-in strategy was proposed in this study.

3.3.2 Zoom-in strategy for improved analysis of multi-composition blended oils

By narrowing the size of the PLS-R model, the zoom-in PLS-R models were developed to improve the analysis of oil compositions at low levels as a follow-up and complimentary step of the full range PLS-R model. Based on the quantitative results of a blended oil sample provided by the full range model, training samples with compositions similar to the measured results were selected from the training set of the full range PLS-R model and applied to establish a new PLS-R model for the blended sample, which was called zoom-in model. For example, for a testing sample containing 95.2% olive oil, 0.6% perilla oil and 4.2% sunflower seed oil as measured by the full range PLS-R model (Table 1, T1), three samples in the whole training set (i.e., olive oil: perilla oil: sunflower seed oil = 100%: 0%: 0%, 90%: 10%: 0% and 90%: 0%: 10%), which had the most similar compositions with the testing sample, were employed to develop the zoom-in model. The established zoom-in model showed a small triangle in the score plot which focused on the testing sample (Figure 2c). Applying this strategy to a quaternary blended oil sample (Table S8, Q2), four training samples were selected to develop the zoom-in model where a mini tetrahedron was observed in the 3D score plot (Figure 2d).

As summarized in Table 1, all the testing samples of the olive oil, perilla oil and sunflower seed oil blends were further quantified by the zoom-in models, and significant improvement in the quantitative results for the low-abundant compositions (~5%) was observed as the ranges of relative errors and precision were narrowed to –22.9-18.3% and 6.6-20.8%, respectively. For the non-existing compositions, the concentrations measured by the zoom-in models were close to 0, preventing erroneous determination of the oil compositions. On the other hand, for the high-abundant compositions, the relative errors and the precision of the results measured by the zoom-in models were in the range of –15.0-14.5% and 0.3-8.0%, respectively, which were quite similar to those obtained by the full range model. The zoom-in strategy was also applied for the quantitative analysis of the olive oil, perilla oil, rice bran oil and sunflower seed oil blends. For compositions of quaternary blended oils, the relative errors of quantitative results by the zoom-in models was typically in the range of –22.4-12.9% (excluding two extreme accuracies –42.5% and 26.1%), and the precision was within 0.4-12.1% (Table S8), which were comparable to those obtained by the full range model. Similarly, the measured results by the zoom-in models were improved for the non-existing compositions, with the highest false positive results significantly decreased from 5.8% to 1.6% (Table S8). Together these results demonstrated that the full range model showed excellent quantitative ability to compositions at medium and higher levels and the zoom-in models could provide better results to low-abundant compositions. Hence, our general strategy to quantify the compositions of blended oils was to use the full range PLS-R model first for screening and then apply the zoom-in PLS-R model for more accurate measurements of the compositions at low levels.

3.4 Comparison of GC-FID and MALDI-MS for quantitative analysis of blended oils

The GC-FID method is a standard method to analyze the FAs contents of edible oils and is commonly employed for authentication of edible oils (Codex, 2015; ISO, 2014). In this study, based on the GC-FID chromatograms (Figure S2) and MALDI-MS spectra obtained by analyzing the same batch of olive oil and sunflower seed oil blends, two optimal PLS-R models (Table S9) were established, and the olive oil concentrations of 5 testing samples were measured by both models for comparison. For samples containing medium and high percentages of olive oil (>30%), both GC-FID- and MALDI-MS-based models provided excellent quantitative results, with relative errors within –1.0-7.8% and –4.1-0.7%, respectively (Table 2). However, for the sample with 7.4% olive oil, the result measured by the MALDI-MS-based model, i.e., 8.0%, was much closer to the actual value than that measured by the GC-FID-based model, i.e., 13.3%.

To further investigate the quantitative analysis of oil compositions at low levels, three prepared samples with 1.0%, 3.2% and 5.2% olive oil were measured by MALDI-MS and GC-FID. The measured concentrations of 0.8%, 2.8% and 5.4%, and 0.1%, 5.0% and 8.1% obtained by the MALDI-MS approach and the GC-FID approach, respectively (Table 2) confirmed the much higher accuracy of the MALDI-MS approach for the compositions at low levels. This was believed to be due to the fact that MALDI-MS directly analyzed the TAGs profiles of oil samples while GC-FID detected the FAs contents converted from the TAGs. For pure olive oil and sunflower seed oil, the cosine correlation scores of their MALDI-MS spectra and GC-FID chromatograms were 0.3068 and 0.5468, respectively, illustrating the variation of the GC-FID chromatograms was smaller than that of the MALDI-MS spectra. The cosine correlation scores between the samples with extremely low olive oil concentrations (i.e., 1.0%, 3.2% and 5.2% olive oil) and the training samples with 20% olive oil were

0.9695, 0.9765 and 0.9805, and 0.9807, 0.9860 and 0.9893 for the MALDI-MS spectra and the GC-FID chromatograms, respectively, indicating larger differences between the MALDI-MS spectra. Therefore, MALDI-MS is more sensitive to slight changes in oil composition than GC-FID and would be more suitable for analyzing compositions at low levels.

GC-FID and MALDI-MS based PLS-R models were also compared for analysis of ternary blends of sunflower seed oil, canola oil and grapeseed oil. For canola oil, the two models showed similar fitting ability and predictive ability, while for sunflower seed oil and grapeseed oil, the MALDI-MS-based model had better fitting ability and improved predictive ability (Table S9). For the sunflower seed oil - canola oil - grapeseed oil blends with blending ratios at 0%:20%:80%, 20%:20%:60%, 50%:10%:40% and 90%:0%:10%, the cosine correlation scores between any two MALDI-MS spectra were lower than those of the corresponding GC-FID chromatograms, which were in the range of 0.9800-0.9980 and 0.9983-0.9997, respectively. Four ternary blended samples with different compositions were measured by the GC-FID and MALDI-MS approaches and the results were shown in Table 2. For samples T3 and T4, comparable results were obtained by the two approaches, while for samples T1 and T2, the MALDI-MS approach provided much better accuracy and precision than the GC-FID approach. As shown in Figure S3, analysis of the VIP values indicated that the quantitative analysis by the GC-FID-based model was mainly determined by the content of docosanoic acid (C22:0) (the only one with VIP values >1). However, C22:0 was very deficient (0-0.7%) in the three pure oils (Table S10), and thus the GC-FID results could be easily affected by random errors. On the other hand, the MALDI-MS-based model had five "important" peaks (VIP value >1) with four of them being the major peaks in the spectra, thus it could provide better quantitative results to different the oil compositions. Compared with the GC-FID approach, there was no need for the MALDI-MS

approach to convert TAGs into FAs, making the quantitative measurements not only simpler but also more accurate due to the maintenance of the delicate TAGs for more sensitive differentiation of complicated oil compositions.

TAGs have similar molecular structures and similar MALDI-MS responses. In this study, the quantitative analysis was based on the relative abundances of TAGs in the MALDI-MS spectra and comparison with data of training samples acquired under the same conditions, minimizing the ion suppression effects. Due to the direct analysis of oil samples in their oily states and much reduced variations of sample spots, the developed MALDI-MS approach offer mass spectra of edible oils with high quality and high reproducibility and is applicable to different types of MALDI-MS equipment (Ng et al., 2018; Ng et al., 2015). Compared with the GC-FID approach, the equipment of the MALDI-MS approach is more expensive, but considering the labor, time and consumables used, the operation cost for analyzing oil samples using MALDI-MS is significantly lower, due to its minimal sample pretreatment and high analysis efficiency, making the MALDI-MS approach robust and economical for analysis of massive samples.

3.5 Quantitation of commercial blended oil products

Previous study demonstrated that specific oil products of the same brands showed highly similar TAGs profiles (Ng, et al., 2018). Therefore, for quantitative analysis of commercial blended oil products, the corresponding pure oil products of the same brands were used to prepare blended oil samples for the establishment of PLS-R models, minimizing the variations of TAGs profiles. Ten commercial blended oil products involving 7 different blending types (Table 3) were analyzed using the MALDI-MS-based PLS-R models (Table S11). Among the eight binary blended products

(products 1-8), the measured concentrations of olive oil, the oil of major concern, for products 4, 5, 7 and 8 were close to the labeled concentrations with relative errors within ±12%, while for products 2, 3 and 6, large differences were observed between the measured and labeled concentrations of olive oil with relative errors exceeding ±20% (Table 3). The concentration of olive oil in product 1 was quite low, which was labeled as 5% and measured as 6.9%, leading to a moderate absolute error (1.9%) but a relatively large relative error (38.0%).

Product 9 was labeled to contain 11% olive oil, 6% flaxseed oil and 83% sunflower seed oil, which were measured as 4.0%, 10.8% and 85.3%, and 2.8%, 6.2% and 91.1%, respectively, by the full range model and the zoom-in model, respectively. The full range and zoom-in models detected sunflower seed oil as the dominant composition, which was consistent with the labeling, but the detected concentrations of olive oil was lower than the labeled one.

Product 10 was labeled to contain 18% olive oil, 15% soybean oil and 67% sunflower seed oil, but both the full range and zoom-in models showed soybean oil as the major composition (~85%) with both olive oil and sunflower seed oil less than 10%. By comparing the MALDI-MS spectra of the pure oils, the manually prepared 10% olive oil – 10% soybean oil – 80% sunflower seed oil blend and product 10, it was obvious that the spectrum of product 10 was quite similar to that of the pure soybean oil, both had the characteristic peak of soybean oil at *m/z* 899.7, which was much higher than that of the manually blended oil (Figure S4). It could be concluded that product 10 was mislabeled and that soybean oil should be the dominant composition of product 10 rather than the labeled sunflower seed oil.

4. Conclusions

A method for rapid and high-throughput quantitation of oil compositions in blended oils has been developed using MALDI-MS and PLS-R. We demonstrated that this method could detect the presence of individual oils at trace level in blended oils, allowed simultaneous measurements of different compositions, and showed excellent quantitative ability for multiple compositions in blended oils particularly with the help of the zoom-in strategy. Compared with the conventional GC-FID approach, the MALDI-MS approach did not need derivatization and column separation, could provide comparable quantitative results and was advantageous for quantitation of the low-abundant compositions. The established method was applied for quantitative analysis of various types of commercial blended oil products, and the results indicated that some commercial products were mislabeled. With the benefits from the simplicity of the sample preparation, the high throughput of MALDI-MS, the robustness of the analytical protocol and the power of PLS-R in data processing, the developed method is very suitable for quantitative analysis of blended oils, especially for those with multiple compositions, and is expected to bring about a significant impact on the quality control of the big blended oil market.

Acknowledgments

Thanks are given to the National Natural Science Foundation of China (Grant No. 81874306) and the Hong Kong Innovation and Technology Fund (Grant No. ITS/070/15) for the financial supports. We also acknowledge the University Research Facility in Life Sciences and the University Research Facility in Chemical and Environmental Analysis of The Hong Kong Polytechnic University are acknowledged for providing the MALDI-TOF/TOF mass spectrometer and for their technical

supports.

References

Alves, J. O., Sena, M. M., & Augusti, R. (2014). Multivariate calibration applied to ESI mass spectrometry data: a tool to quantify adulteration in extra virgin olive oil with inexpensive edible oils. *Analytical Methods, 6*(18), 7502-7509. https://doi.org/10.1039/c4ay00967c.

Ayorinde, F. O., Elhilo, E., & Hlongwane, C. (1999). Matrix-assisted laser desorption/ionization tme-of-flight mass spectrometry of canola, castor and olive oils. *Rapid Communications in Mass Spectrometry, 13*, 737–739. https://doi.org/10.1002/(SICI)1097-0231(19990430)13:8<737::AID-RCM552>3.0.CO;2-L.

Calvano, C. D., Palmisano, F., & Zambonin, C. G. (2005). Laser desorption/ionization time-of-flight mass spectrometry of triacylglycerols in oils. *Rapid Communications in Mass Spectrometry, 19*(10), 1315-1320. https://doi.org/10.1002/rcm.1933.

Cho, S. K., Yang, S. O., Kim, S. H., Kim, H., Ko, J. S., Riu, K. Z., . . . Choi, H. K. (2009). Classification and prediction of free-radical scavenging activities of dangyuja (Citrus grandis Osbeck) fruit extracts using (1)H NMR spectroscopy and multivariate statistical analysis. *Journal of Pharmaceutical and Biomedical Analysis, 49*(2), 567-571. https://doi.org/10.1016/j.jpba.2008.11.037.

Choi, H., Lee, E., & Lee, K. G. (2014). Quality Evaluation of Noble Mixed Oil Blended with Palm and Canola Oil. *Journal of Oleo Science, 63*(7), 653-660. https://doi.org/10.5650/jos.ess14023.

Choudhary, M., Grover, K., & Kaur, G. (2015). Development of rice bran oil blends for quality improvement. *Food Chemistry, 173*, 770-777. https://doi.org/10.1016/j.foodchem.2014.10.051.

Codex. (2015). CODEX standard for named vegetable oils. (Vol. CODEX STAN 210-1999): Codex Alimentarius.

Cozzolino, R., & De Giulio, B. (2011). Application of ESI and MALDI-TOF MS for triacylglycerols analysis in edible oils. *European Journal of Lipid Science and Technology, 113*(2), 160-167. https://doi.org/10.1002/ejlt.201000429.

de la Mata-Espinosa, P., Bosque-Sendra, J. M., Bro, R., & Cuadros-Rodríguez, L. (2011). Olive oil quantification of edible vegetable oil blends using triacylglycerols chromatographic fingerprints and chemometric tools. *Talanta, 85*(1), 177-182. https://doi.org/10.1016/j.talanta.2011.03.049.

EC. (2012). Commission implementing regulation (EC) No 29/2012 of 13 January 2013 on marketing standards for olive oil (codification). (Vol. No 29/2012): Official Journal of the European Union.

El-Abassy, R. M., Donfack, P., & Materny, A. (2009). Visible Raman spectroscopy for the discrimination of olive oils from different vegetable oils and the detection of adulteration.

Journal of Raman Spectroscopy, 40(9), 1284-1289. https://doi.org/10.1002/jrs.2279.

FAO. (2010). Summary of conclusions and dietary recommendations on total fat and fatty acids. In *Fats and fatty aicds in human nutritions: Report of an expert consultation* (pp. 9-19). Rome: Food and Agriculture Organization of the United Nations.

Fasciotti, M., & Pereira Netto, A. D. (2010). Optimization and application of methods of triacylglycerol evaluation for characterization of olive oil adulteration by soybean oil with HPLC-APCI-MS-MS. *Talanta, 81*(3), 1116-1125. https://doi.org/10.1016/j.talanta.2010.02.006.

FSSAI. (2011). Food safety and standards (packaging and labelling) regulations, 2011. Food Safety and Standards Authority of India.

ISO. (2014). Animal and vegetable fats and oils - Gas chromatography of fatty acid methyl esters. (Vol. ISO 12966): International Organization for Standardization.

Jović, O., Smolić, T., Primožič, I., & Hrenar, T. (2016). Spectroscopic and Chemometric Analysis of Binary and Ternary Edible Oil Mixtures: Qualitative and Quantitative Study. *Analytical Chemistry, 88*(8), 4516-4524. https://doi.org/10.1021/acs.analchem.6b00505.

Kuo, T. H., Kuei, M. S., Hsiao, Y., Chung, H. H., Hsu, C. C., & Chen, H. J. (2019). Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry Typings of Edible Oils through Spectral Networking of Triacylglycerol Fingerprints. *ACS Omega, 4*(13), 15734-15741. https://doi.org/10.1021/acsomega.9b02433.

Lay, J. O. J., Liyanage, R., Durham, B., & Brooks, J. (2006). Rapid characterization of edible oils by direct matrix-assisted laser desorption/ionization time-of-flight mass spectrometry analysis using triacylglycerols. *Rapid Communication in Mass Spectrometry, 20*(6), 952-958. https://doi.org/10.1002/rcm.2394.

Li, B., Morris, J., & Martin, E. B. (2002). Model selection for partial least squares regression. *Chemometrics and Intelligent Laboratory Systems, 64*, 79–89. https://doi.org/10.1016/S0169-7439(02)00051-5.

Li, B., Wang, H., Zhao, Q., Ouyang, J., & Wu, Y. (2015). Rapid detection of authenticity and adulteration of walnut oil by FTIR and fluorescence spectroscopy: a comparative study. *Food Chemistry, 181*, 25-30. https://doi.org/10.1016/j.foodchem.2015.02.079.

Milanez, K. D. T. M., Nóbrega, T. C. A., Nascimento, D. S., Insausti, M., Band, B. S. F., & Pontes, M. J. C. (2017). Multivariate modeling for detecting adulteration of extra virgin olive oil with soybean oil using fluorescence and UV–Vis spectroscopies: A preliminary approach. *LWT - Food Science and Technology, 85*, 9-15. https://doi.org/10.1016/j.lwt.2017.06.060.

Monfreda, M., Gobbi, L., & Grippa, A. (2014). Blends of olive oil and seeds oils: characterisation and olive oil quantification using fatty acids composition and chemometric tools. Part II. *Food Chemistry, 145*, 584-592. https://doi.org/10.1016/j.foodchem.2013.07.141.

Ng, T. T., Li, S., Ng, C. C. A., So, P. K., Wong, T. F., Li, Z. Y., . . . Yao, Z. P. (2018). Establishment of a spectral database for classification of edible oils using matrix-assisted laser desorption/ionization mass spectrometry. *Food Chemistry, 252*, 335-342. https://doi.org/10.1016/j.foodchem.2018.01.125.

Ng, T. T., So, P. K., Zheng, B., & Yao, Z. P. (2015). Rapid screening of mixed edible oils and gutter oils by matrix-assisted laser desorption/ionization mass spectrometry. *Analytica Chimica Acta, 884*, 70-76. https://doi.org/10.1016/j.aca.2015.05.013.

NHC, & SAMR. (2018). National food safety standard - vegetable oil. (Vol. GB 2716-2018): National Health Commission & State Administration for Market Regulation.

OECD/FAO. (2018). Oilseeds and oilseed products. In *OECD-FAO Agricultural Outlook 2018-2027* (pp. 127-138). Paris: The Organisation for Economic Co-operation and Development (OECD) publishing.

Olivieri, A. C., Faber, N. M., Ferré, J., Boqué, R., Kalivas, J. H., & Mark, H. (2006). Uncertainty estimation and figures of merit for multivariate calibration (IUPAC Technical Report). *Pure and Applied Chemistry, 78*(3), 633-661. https://doi.org/10.1351/pac200678030633.

Ortiz, M. C., Sarabia, L. A., Herrero, A., Sánchez, M. S., Sanz, M. B., Rueda, M. E., . . . Meléndez, M. E. (2003). Capability of detection of an analytical method evaluating false positive and false negative (ISO 11843) with partial least squares. *Chemometrics and Intelligent Laboratory Systems, 69*(1-2), 21-33. https://doi.org/10.1016/s0169-7439(03)00110-2.

Park, Y. W., Chang, P.-S., & Lee, J. (2010). Application of triacylglycerol and fatty acid analyses to discriminate blended sesame oil with soybean oil. *Food Chemistry, 123*(2), 377-383. https://doi.org/10.1016/j.foodchem.2010.04.049.

Poulli, K. I., Mousdis, G. A., & Georgiou, C. A. (2006). Synchronous fluorescence spectroscopy for quantitative determination of virgin olive oil adulteration with sunflower oil. *Analytical and Bioanalytical Chemistry, 386*(5), 1571-1575. https://doi.org/10.1007/s00216-006-0729-2.

Rohman, A., & Che Man, Y. B. (2010). Potential Use of FTIR-ATR Spectroscopic Method for Determination of Virgin Coconut Oil and Extra Virgin Olive Oil in Ternary Mixture Systems. *Food Analytical Methods, 4*(2), 155-162. https://doi.org/10.1007/s12161-010-9156-2.

Rohman, A., & Che Man, Y. B. (2011). Determination of extra virgin olive oil in quaternary mixture using FTIR spectroscopy and multivariate calibration. *Spectroscopy, 26*(3), 203-211. https://doi.org/10.3233/SPE-2011-0538.

Rohman, A., & Che Man, Y. B. (2012). Authentication of Extra Virgin Olive Oil from Sesame Oil Using FTIR Spectroscopy and Gas Chromatography. *International Journal of Food Properties, 15*(6), 1309-1318. https://doi.org/10.1080/10942912.2010.521607.

Roxborough, S. (2018). Olive oil products fraud uncovered in China. Retrieved from: https://www.oliveoiltimes.com/olive-oil-business/asia/olive-oil-products-fraud-uncovered-in-china/62774 Accessed 05 April 2020.

Ruiz-Samblas, C., Marini, F., Cuadros-Rodriguez, L., & Gonzalez-Casado, A. (2012). Quantification of blending of olive oils and edible vegetable oils by triacylglycerol fingerprint gas chromatography and chemometric tools. *Journal of Chromatography B, 910*, 71-77. https://doi.org/10.1016/j.jchromb.2012.01.026.

SCHILLER, J., SÜß, R., PETKOVI, M., & ARNOLD, K. (2002). Triacylglycerol analysis of vegetable oils by matrix ‐ assisted laser desorption and ionization time ‐ of ‐ flight (MALDI ‐ TOF) mass spectrometry and 31P NMR spectroscopy. *Journal of Food Lipids, 9*, 185-200. https://doi.org/10.1111/j.1745-4522.2002.tb00218.x.

Tien, Y. P., & Hsu, E. (2017). Ting Hsin Former Chairman Jailed in Oil Scandal. Retrieved from: http://focustaiwan.tw/news/asoc/201707280011.aspx Accessed 05 April 2020.

Veerasamy, R., Rajak, H., Jain, A., Sivadasan, S., Varghese, C. P., & Agrawal, R. K. (2011). Validation of QSAR Models - Strategies and Importance. *International Journal of Drug Design and Discovery, 2*(3), 511-519.

Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems, 58*, 109–130. https://doi.org/10.1016/S0169-7439(01)00155-1.

Xie, J., Liu, T., Yu, Y., Song, G., & Hu, Y. (2013). Rapid Detection and Quantification by GC–MS of Camellia Seed Oil Adulterated with Soybean Oil. *Journal of the American Oil Chemists' Society, 90*(5), 641-646. https://doi.org/10.1007/s11746-013-2209-0.

Table 1. Quantitative results of olive oil (OO) – perilla oil (PR) – sunflower seed oil (SF) blends by the full range and zoom-in models

Sample	Oil species	Actual con. (%)	Full range model			Zoom-in model		
			Measured con. (%)	Relative error (%)	RSD (%)	Measured con. (%)	Relative error (%)	RSD (%)
T1	OO	100.0	95.2±1.0	–4.8	1.1	99.7±0.3	–0.3	0.3
	PR	0.0	0.6±1.4	/	/	–0.4±0.4	/	/
	SF	0.0	4.2±0.8	/	/	0.7±0.5	/	/
T2	OO	0.0	3.8±0.3	/	/	–0.2±0.4	/	/
	PR	100.0	97.4±1.0	–2.7	1.0	99.8±0.7	–0.2	0.7
	SF	0.0	–1.4±1.2	/	/	0.4±0.6	/	/
T3	OO	0.0	–0.3±0.6	/	/	–1.3±0.6	/	/
	PR	0.0	7.4±1.4	/	/	0.5±0.5	/	/
	SF	100.0	92.6±0.5	–7.4	0.5	99.6±0.7	–0.4	0.7
T4	OO	0.0	1.5±0.5	/	/	0.4±0.4	/	/
	PR	49.8	52.7±1.0	5.9	2.0	52.7±0.8	5.8	1.5
	SF	50.2	45.8±0.7	–8.8	1.6	46.9±0.6	–6.6	1.2
T5	OO	49.8	48.0±0.7	–3.5	1.4	47.3±0.4	–5.0	0.9
	PR	0.0	2.9±0.6	/	/	1.0±0.5	/	/
	SF	50.2	49.4±1.4	–1.5	2.9	51.4±0.5	2.5	1.0
T6	OO	50.2	44.4±0.4	–11.5	1.0	46.3±0.3	–7.7	0.7
	PR	49.8	55.8±1.0	12.0	1.9	53.8±0.5	8.0	1.0
	SF	0.0	–0.2±0.7	/	/	–0.1±0.6	/	/
T7	OO	5.0	4.2±0.7	–15.2	17.6	3.8±0.3	–22.9	8.4
	PR	5.4	7.6±1.5	41.9	19.2	6.4±0.6	18.3	9.0
	SF	89.6	88.2±0.9	–1.7	1.0	90.1±0.8	0.5	0.9
T8	OO	5.0	5.4±0.1	7.7	2.6	4.4±0.3	–12.7	6.6
	PR	89.8	91.7±1.2	2.0	1.3	90.6±0.8	0.8	0.8
	SF	5.1	2.8±1.1	–45.0	38.9	5.0±0.7	–2.3	14.5
T9	OO	89.7	88.7±0.7	–1.1	0.8	89.8±1.1	0.2	1.2
	PR	5.0	4.6±0.8	–8.8	18.5	4.4±0.9	–13.0	20.8
	SF	5.3	6.8±0.8	28.9	12.4	5.8±0.7	9.2	11.4
T10	OO	20.2	18.0±0.7	–10.7	4.1	18.6±1.5	–7.5	8.0
	PR	20.1	21.3±1.0	5.9	4.7	20.6±0.8	2.4	3.9
	SF	59.7	60.7±1.0	1.6	1.6	60.7±0.8	1.7	1.2
T11	OO	20.1	16.7±0.4	–16.7	2.1	17.1±0.3	–15.0	1.6
	PR	59.8	66.1±0.6	10.5	1.0	64.0±0.4	6.9	0.6
	SF	20.1	17.1±0.6	–14.6	3.5	19.0±0.5	–5.5	2.8
T12	OO	59.7	60.2±0.5	0.8	0.9	56.8±0.5	–4.8	0.9
	PR	20.2	19.4±0.2	–3.8	0.9	23.1±0.2	14.5	0.9
	SF	20.1	20.4±0.5	1.3	2.3	20.1±0.4	–0.3	1.9

Table 2. Comparison of PLS-R quantitative results of olive oil (OO) – sunflower seed oil (SF) binary blends (B1-B8) and sunflower seed oil (SF) – canola oil (CA) – grapeseed oil (GP) ternary blends (T1-T4) based on the GC-FID chromatograms and MALDI-MS spectra

Sample	Oil species	Actual con. (%)	GC-FID			MALDI-MS		
			Measured con. (%)	Relative error (%)	RSD (%)	Measured con. (%)	Relative error (%)	RSD (%)
B1	OO	1.0	0.1±0.8	–93.5	1173.3	0.8±0.8	–20.0	95.5
B2	OO	3.2	5.0±0.1	56.8	1.7	2.8±0.6	–13.8	20.5
B3	OO	5.2	8.1±0.7	56.4	8.5	5.4±2.1	3.8	38.8
B4	OO	7.4	13.3±0.4	79.5	2.8	8.0±0.7	7.7	9.2
B5	OO	30.1	32.4±0.6	7.8	1.8	28.9±0.8	–4.1	2.8
B6	OO	49.9	51.5±0.5	3.1	1.0	49.7±0.6	–0.5	1.2
B7	OO	69.9	70.4±0.6	0.6	0.8	70.4±0.7	0.7	0.9
B8	OO	92.3	91.3±0.1	–1.0	0.1	92.7±0.5	0.5	0.5
T1	SF	0.0	16.3±9.9	/	/	–0.6±3.2	/	/
	CA	23.0	22.4±1.1	–2.3	5.1	23.4±1.4	1.6	5.8
	GP	77.0	61.2±9.1	–20.5	14.8	77.2±3.3	0.2	4.2
T2	SF	20.0	25.9±9.9	29.6	38.0	18.5±2.5	–7.6	13.5
	CA	18.0	19.4±2.4	8.1	12.5	18.0±0.8	–0.2	4.4
	GP	62.0	54.6±7.5	–11.9	13.8	63.6±2.4	2.5	3.8
T3	SF	50.0	52.4±0.8	4.8	1.6	47.1±2.2	–5.7	4.6
	CA	13.1	14.1±1.4	8.3	9.6	13.5±0.9	3.6	6.8
	GP	37.0	33.5±0.8	–9.4	2.5	39.4±2.5	6.4	6.2
T4	SF	89.8	89.7±1.6	–0.2	1.8	86.8±3.0	–3.4	3.5
	CA	4.1	5.2±1.7	25.8	33.3	5.1±1.5	24.1	29.4
	GP	6.0	5.1±2.6	–15.2	50.8	8.1±2.6	34.0	32.4

Table 3. Quantitative results of commercial blended oil products

Product	Type[a]	Model[b]	Oil species	Labeled con. (%)	Model range	Measured con. (%)	Relative error (%)	RSD (%)
1	OO-SF	M1	OO	5	Full	6.9±1.2	38.0	16.9
2	OO-SF	M2	OO	10	Full	3.5±0.5	–65.0	15.3
3	OO-SF		OO	20	Full	12.9±0.3	–35.5	2.5
4	OO-CA	M3	OO	10	Full	9.4±1.7	–6.0	18.2
5	OO-CO	M4	OO	6	Full	6.5±0.6	8.3	9.6
6	OO-CO	M5	OO	20	Full	15.6±1.8	–22.0	11.8
7	OO-PA	M6	OO	50	Full	45.0±3.0	–10.0	6.7
8	OO-PR	M7	OO	50	Full	55.6±1.2	11.2	2.1
9	OO-FS-SF	M8	OO	11	Full	4.0±0.5	–63.6	12.3
					Zoom-in	2.8±0.4	–74.5	16.3
			FS	6	Full	10.8±0.4	80.0	4.2
					Zoom-in	6.2±0.6	3.3	9.4
			SF	83	Full	85.3±0.4	2.8	0.4
					Zoom-in	91.1±0.6	9.8	0.7
10	OO-SO-SF	M9	OO	18	Full	4.3±0.4	–76.1	9.2
					Zoom-in	6.8±0.9	–62.2	13.2
			SO	15	Full	85.9±1.0	472.7	1.2
					Zoom-in	84.4±1.9	462.7	2.2
			SF	67	Full	9.8±1.2	–85.4	12.3
					Zoom-in	9.5±0.9	–85.8	9.6

[a] OO: olive oil, SF: sunflower seed oil, CA: canola oil, CO: corn oil, PA: high oleic peanut oil, PR: perilla oil, FS: flaxseed oil, SO: soybean oil.

[b] See Table S11 for the information of the corresponding PLS-R models.

Figure Captions

Figure 1. The TAGs region of the MALDI-MS spectra for (a) 100% sunflower seed oil, (b) 40% olive oil – 60% sunflower seed oil blend, (c) 60% olive oil – 40% sunflower seed oil blend, and (d) 100% olive oil, with the identities of the major peaks assigned. P, palmitic acid; L, linoleic acid; O, oleic acid; S, steric acid.

Figure 2. The score plots of the (a) full range and (c) zoom-in PLS-R models of olive oil (O), perilla oil (P) and sunflower seed oil (S) blends, and the 3D score plots of the (b) full range and (d) zoom-in PLS-R models of olive oil (O), perilla oil (P), rice bran oil (R) and sunflower seed oil (S) blends. Axes t[1], t[2] and t[3] represent the X-scores of the first, second and third PLS components, respectively.

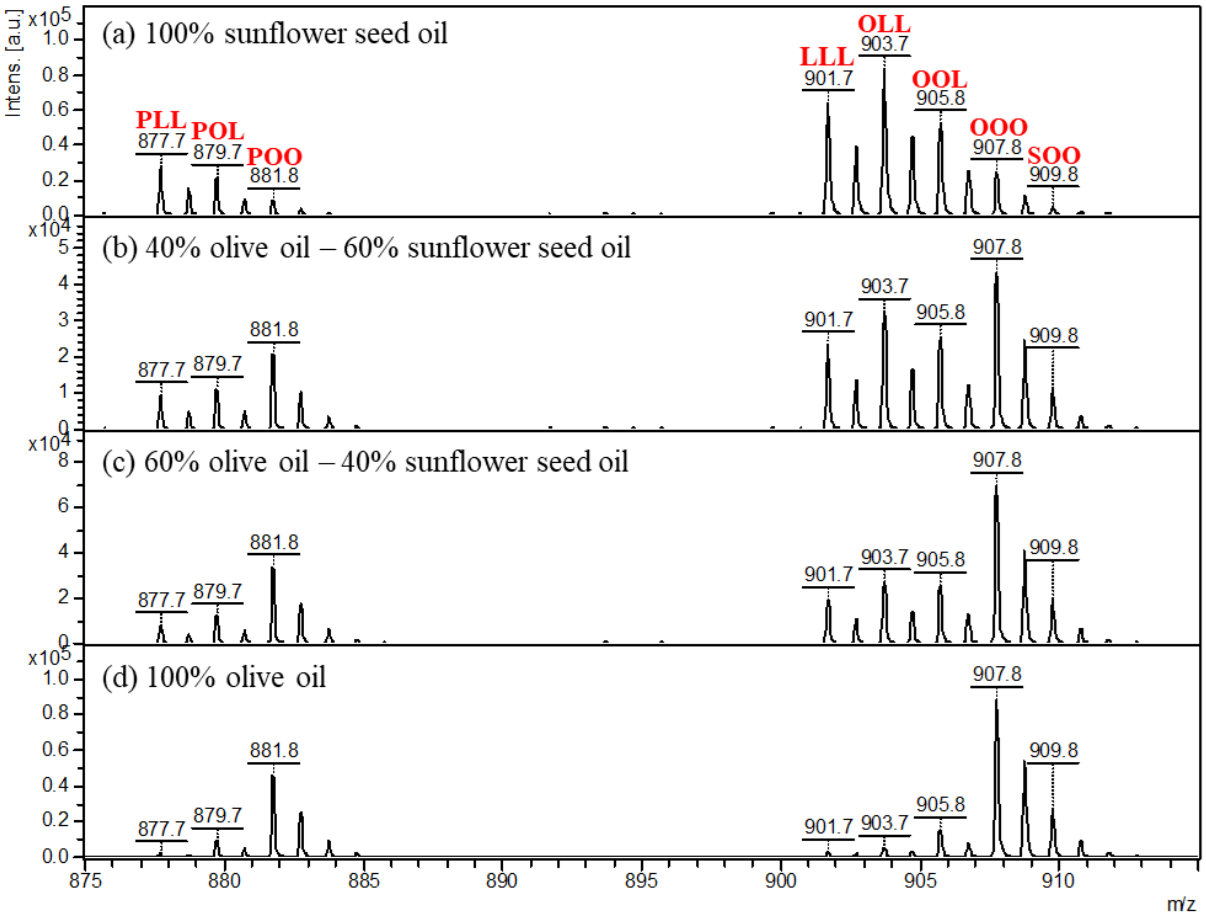
Intens. [a.u.]
(a) 100% sunflower seed oil
PLL
POL
POO
OLL
LLL
OOL
OOO
SOO
877.7
879.7
881.8
901.7
903.7
905.8
907.8
909.8
(b) 40% olive oil – 60% sunflower seed oil
(c) 60% olive oil – 40% sunflower seed oil
(d) 100% olive oil
m/z

Figure 1.

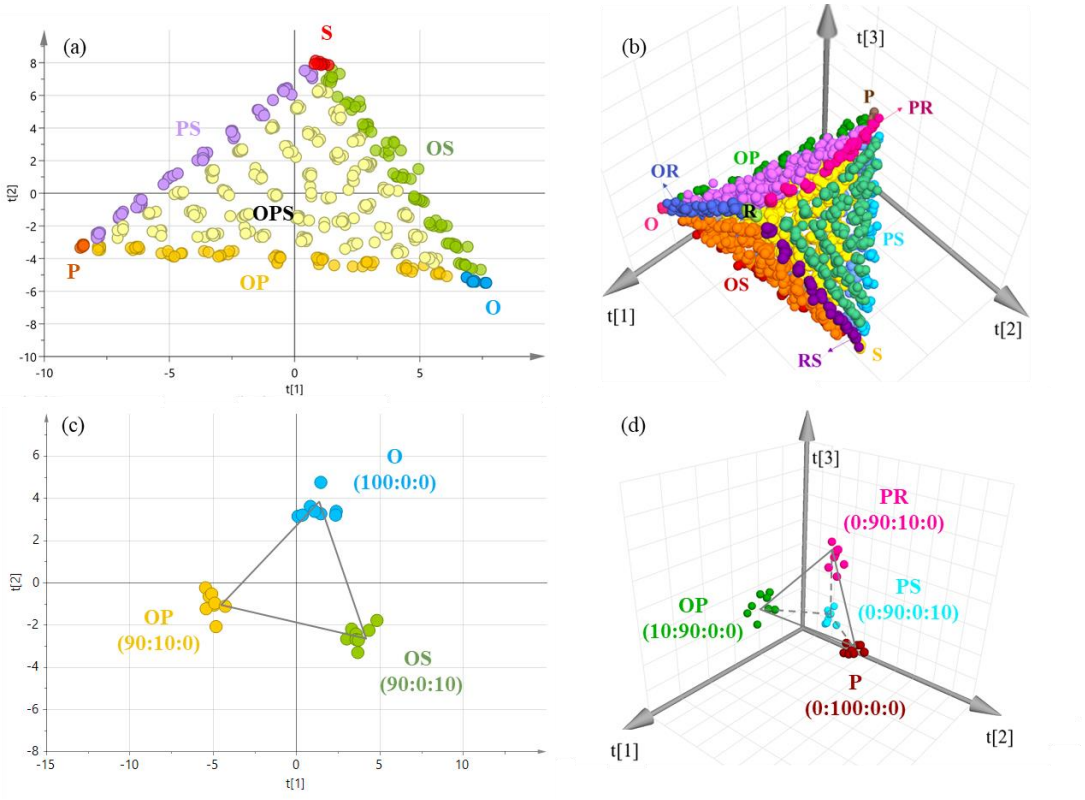
(a)
S
PS
OS
OPS
P
OP
O
t[1]
t[2]
(b)
t[3]
P
PR
OR
OP
O
R
PS
OS
RS
S
t[1]
t[2]
(c)
O
(100:0:0)
OP
(90:10:0)
OS
(90:0:10)
t[1]
t[2]
(d)
t[3]
PR
(0:90:10:0)
PS
(0:90:0:10)
OP
(10:90:0:0)
P
(0:100:0:0)
t[1]
t[2]

Figure 2.

Supporting information

Quantitative Analysis of Blended Oils by Matrix-assisted Laser Desorption/Ionization Mass Spectrometry and Partial Least Squares Regression

Suying Li,[a,b] Tsz-Tsun Ng[a,b] and Zhong-Ping Yao[a,b,*]

[a]State Key Laboratory of Chemical Biology and Drug Discovery, Food Safety and Technology Research Centre and Department of Applied Biology and Chemical Technology, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong Special Administrative Region, China

[b]State Key Laboratory of Chinese Medicine and Molecular Pharmacology (Incubation) and Shenzhen Key Laboratory of Food Biological Safety Control, Shenzhen Research Institute of Hong Kong Polytechnic University, Shenzhen 518057, China

*To whom correspondence should be addressed:

Dr. Zhong-Ping Yao
Department of Applied Biology and Chemical Technology
The Hong Kong Polytechnic University
Hung Hom, Kowloon
Hong Kong
Tel: (852) 34008792. Fax: (852)23649932.
E-mail: zhongping.yao@polyu.edu.hk

Contents:

Additional Experimental Details

S1. Determination of optimal A

Generally, increasing the number of PLS components would result in a better fitting of the PLS-R model to the original data but might decrease its predictive ability due to noise information, which is called over-fitting. Wold's R criterion and RMSE criterion based on the results of cross-validation were applied to determine the optimal number of PLS components (A) and prevent the over-fitting. Wold's R criterion calculates the ratio of the predicted error sum of squares (PRESS) for (*m + 1*) PLS components and *m* PLS components, which is equal to the square of RMSEcv ratio for (*m + 1*) PLS components and *m* PLS components as shown in equation (1). When R exceeds threshold (0.90 or 0.95), an optimal A is obtained as *m*. For RMSE criterion, the root mean square error (RMSE) is calculated as in equation (2), and the optimal A is obtained with minimal RMSE.

$$\mathrm{PRESS} = \sum_{i=1}^{N} (y_{i,actual} - y_{i,measured})^2$$

$$\mathrm{RMSEcv} = \sqrt{\frac{\sum_{i=1}^{N} (y_{i,actual} - y_{i,measured})^2}{N}}$$

$$\mathrm{R} = \frac{\mathrm{PRESS}(m+1)}{\mathrm{PRESS}(m)} = \left(\frac{\mathrm{RMSEcv}(m+1)}{\mathrm{RMSEcv}(m)}\right)^2 \quad (1)$$

$$\mathrm{RMSE} = \mathrm{RMSEcv} \times \sqrt{\frac{N}{N-A-1}} \quad (2)$$

N is the number of data in the training set, and $y_{i,\ measured}$ and $y_{i,\ actual}$ are the measured concentrations and actual concentrations of samples, respectively.

S2. Determination of limit of detection

The limit of detection (LOD) of the PLS-R model could be calculated using equation (3) (Ortiz et al., 2003),

$$\mathrm{LOD}_{cal} = \Delta(\alpha, \beta)\frac{\sigma}{b}\sqrt{\frac{1}{K} + \frac{1}{N} + \frac{(\bar{y}_{actual})^2}{\sum_{i=1}^{N}\left(y_{i,actual} - \bar{y}_{actual}\right)^2}} \quad (3)$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(y_{i,measured} - y_{i,actual})^2}{N-2}}$$

where $\Delta(\alpha, \beta)$ is the non-centrality parameter of non-central Student's t-distribution with (n-2) degrees of freedom, n is the number of training samples, N is the number of data in the training set, σ is the standard residual deviation of regression curve by plotting the measured concentrations ($y_{i,\,measured}$) of samples in the training set against their actual concentrations ($y_{i,\,actual}$), b is the slope of the regression curve, K is the number of determination performed on each sample, and $\bar{y}_{actual}$ is the mean actual concentration of all the samples in the training set.

For the PLS-R model of olive oil and sunflower seed oil blends, the degree of freedom was 4 and the $\Delta(\alpha, \beta)$ with confidence level of 95% was 4.07 (Clayton, Hines, & Elkins, 1987). The slope and the standard residual deviation of the regression curve between the measured olive oil concentrations and the actual olive oil concentrations of training samples were 1.0 and 0.5%, respectively. Therefore, the LOD of olive oil was calculated as 0.9%. Furthermore, a blank sample (i.e., the pure sunflower seed oil) was measured for 20 replicates, and the olive oil concentration measured by the established PLS-R model was -0.4 ± 0.4%. As IUPAC recommended, to make a correct positive detection decision with sufficiently high probability, the probabilities of both false positive and false negative should be considered (Olivieri, Faber, Ferré, Boqué, Kalivas, & Mark, 2006). Thus, the LOD based on the measured results of the blank sample was calculated also as 0.9%, based on equation (4).

$$\text{LOD} = \text{Average (blank)} + 3.3 \times \text{Standard deviation} \quad (4)$$

S3. Similarity Scoring Method

Similarity between two MALDI-MS spectra was determined by cosine correlation (Tabb, MacCoss, Wu, Anderson, & Yates, 2003), as defined below,

$$cos\ \theta = \frac{\sum i_A i_B}{\sqrt{\sum i_A^2 \sum i_B^2}}$$

where θ is the spectral contrast angle between the MALDI-MS spectra of selected oil sample A and sample B, i_A is the relative intensities of peaks from spectrum A and i_B is the relative intensities of peak from spectrum B. For the first sum, only when a peak at a particular *m/z* is observed in both spectrum A and spectrum B would the relative intensities of this peak in both spectra be multiplied together, otherwise the product of i_A multiplied by i_B should be zero. The calculation of similarity of the GC-FID chromatograms between selected oil sample A and sample B wad the same as above.

References

Clayton, C. A., Hines, J. W., & Elkins, P. D. (1987). Detection Limits with Specified Assurance Probabilities. *Analytical Chemistry, 59*, 2506-2514. https://doi.org/10.1021/ac00147a014.

Olivieri, A. C., Faber, N. M., Ferré, J., Boqué, R., Kalivas, J. H., & Mark, H. (2006). Uncertainty estimation and figures of merit for multivariate calibration (IUPAC Technical Report). *Pure and Applied Chemistry, 78*(3), 633-661. https://doi.org/10.1351/pac200678030633.

Ortiz, M. C., Sarabia, L. A., Herrero, A., Sánchez, M. S., Sanz, M. B., Rueda, M. E., . . . Meléndez, M. E. (2003). Capability of detection of an analytical method evaluating false positive and false negative (ISO 11843) with partial least squares. *Chemometrics and Intelligent Laboratory Systems, 69*(1-2), 21-33. https://doi.org/10.1016/s0169-7439(03)00110-2.

Tabb, D. L., MacCoss, M. J., Wu, C. C., Anderson, S. D., & Yates, J. R. (2003). Similarity among Tandem Mass Spectra from Proteomic Experiments Detection, Significance, and Utility. *Analytical Chemistry, 75*, 2470-2477. https://doi.org/10.1021/ac026424o.

Table S1. Vegetable oil products collected from the market

No.	Type	Brand	Origin
1	Olive oil	A	Hong Kong
2	Sunflower seed oil	A	Hong Kong
3	Rice bran oil	B	Hong Kong
4	Grapeseed oil	B	Hong Kong
5	Perilla oil	C	Hong Kong
6	Olive oil	A	Mainland China
7	Sunflower seed oil	A	Mainland China
8	Canola oil	A	Mainland China
9	Olive oil	D	Mainland China
10	Sunflower seed oil	D	Mainland China
11	Corn oil	D	Mainland China
12	Olive oil	E	Mainland China
13	Corn oil	E	Mainland China
14	Olive oil	F	Mainland China
15	High oleic acid peanut oil	F	Mainland China
16	Olive oil	G	Mainland China
17	Perilla oil	G	Mainland China
18	Olive oil	H	Mainland China
19	Soybean oil	H	Mainland China
20	Sunflower seed oil	H	Mainland China
21	Olive oil	I	Mainland China
22	Flaxseed oil	I	Mainland China
23	Sunflower seed oil	I	Mainland China
24	Olive oil - sunflower seed oil blend	A	Mainland China
25	Olive oil - sunflower seed oil blend	D	Mainland China
26	Olive oil - sunflower seed oil blend	D	Mainland China
27	Olive oil - canola oil blend	A	Mainland China
28	Olive oil - corn oil blend	D	Mainland China
29	Olive oil - corn oil blend	E	Mainland China
30	Olive oil - perilla oil blend	G	Mainland China
31	Olive oil - high oleic acid peanut oil blend	F	Mainland China
32	Olive oil - soybean oil - sunflower seed oil blend	H	Mainland China
33	Olive oil - flaxseed oil - sunflower seed oil blend	I	Mainland China

Table S2-1. Blended oil samples prepared for establishment of PLS-R models for binary and ternary blended oils

<table>
<tr><td rowspan="5">Binary blended oils</td><td colspan="3">Training set (6)</td><td>Testing set (5)</td></tr>
<tr><td colspan="3">Pure (2)</td><td rowspan="4">8%:92%
30%:70%
50%:50%
70%:30%
92%:8%</td></tr>
<tr><td colspan="2">0%:100%</td><td>100%:0%</td></tr>
<tr><td colspan="3">Binary blends (4)</td></tr>
<tr><td colspan="2">20%:80%
40%:60%</td><td>60%:40%
80%:20%</td></tr>
<tr><td rowspan="8">Ternary blended oils</td><td colspan="3">Training set (66)</td><td>Testing set (12)</td></tr>
<tr><td colspan="3">Pure (3)</td><td rowspan="7">50%:50%:0%
50%:0%:50%
0%:50%:50%
5%:5%:90%
5%:90%:5%
90%:5%:5%
20%:20%:60%
20%:60%:20%
60%:20%:20%
0%:0%:100%
0%:100%:0%
100%:0%:0%</td></tr>
<tr><td>0%:0%:100%</td><td>0%:100%:0%</td><td>100%:0%:0%</td></tr>
<tr><td colspan="3">Binary blends (9×3)</td></tr>
<tr><td>0%:10%:90%
0%:20%:80%
0%:30%:70%
0%:40%:60%
0%:50%:50%
0%:60%:40%
0%:70%:30%
0%:80%:20%
0%:90%:10%</td><td>10%:90%:0%
20%:80%:0%
30%:70%:0%
40%:60%:0%
50%:50%:0%
60%:40%:0%
70%:30%:0%
80%:20%:0%
90%:10%:0%</td><td>10%:0%:90%
20%:0%:80%
30%:0%:70%
40%:0%:60%
50%:0%:50%
60%:0%:40%
70%:0%:30%
80%:0%:20%
90%:0%:10%</td></tr>
<tr><td colspan="3">Ternary blends (36)</td></tr>
<tr><td>10%:10%:80%
10%:20%:70%
10%:30%:60%
10%:40%:50%
10%:50%:40%
10%:60%:30%
10%:70%:20%
10%:80%:10%
20%:10%:70%
20%:20%:60%
20%:30%:50%
20%:40%:40%</td><td>20%:50%:30%
20%:60%:20%
20%:70%:10%
30%:10%:60%
30%:20%:50%
30%:30%:40%
30%:40%:30%
30%:50%:20%
30%:60%:10%
40%:10%:50%
40%:20%:40%
40%:30%:30%</td><td>40%:40%:20%
40%:50%:10%
50%:10%:40%
50%:20%:30%
50%:30%:20%
50%:40%:10%
60%:10%:30%
60%:20%:20%
60%:30%:10%
70%:10%:20%
70%:20%:10%
80%:10%:10%</td></tr>
</table>

Table S2-2 Blended oil samples prepared for establishment of PLS-R models for quaternary blended oils

<table>
<tr><td rowspan="15">Quaternary blended oils</td><td colspan="4">Training set (286)</td><td>Testing set (13)</td></tr>
<tr><td colspan="4">Pure (4)</td><td rowspan="15">10%:10%:10%:70%
10%:10%:70%:10%
10%:70%:10%:10%
70%:10%:10%:10%
20%:20%:20%:40%
20%:20%:40%:20%
20%:40%:20%:20%
40%:20%:20%:20%
25%:25%:25%:25%
0%:0%:0%:100%
0%:0%:100%:0%
0%:100%:0%:0%
100%:0%:0%:0%</td></tr>
<tr><td>0%:0%:0%:100%</td><td>0%:0%:100%:0%</td><td>0%:100%:0%:0%</td><td>100%:0%:0%:0%</td></tr>
<tr><td colspan="4">Binary blends (9×6)</td></tr>
<tr><td>10%:0%:0%:90%
20%:0%:0%:80%
30%:0%:0%:70%
…
90%:0%:0%:10%</td><td colspan="2">10%:0%:90%:0%
20%:0%:80%:0%
30%:0%:70%:0%
…
90%:0%:10%:0%</td><td>10%:90%:0%:0%
20%:80%:0%:0%
30%:70%:0%:0%
…
90%:10%:0%:0%</td></tr>
<tr><td colspan="4"></td></tr>
<tr><td>0%:10%:90%:0%
0%:20%:80%:0%
0%:30%:70%:0%
…
0%:90%:10%:0%</td><td colspan="2">0%:10%:0%:90%
0%:20%:0%:80%
0%:30%:0%:70%
…
0%:90%:0%:10%</td><td>0%:0%:10%:90%
0%:0%:20%:80%
0%:0%:30%:70%
…
0%:0%:90%:10%</td></tr>
<tr><td colspan="4">Ternary blends (36×4)</td></tr>
<tr><td>10%:10%:80%:0%
10%:20%:70%:0%
10%:30%:60%:0%
…
80%:10%:10%:0%</td><td>10%:0%:10%:80%
10%:0%:20%:70%
10%:0%:30%:60%
…
80%:0%:10%:10%</td><td>10%:10%:0%:80%
10%:20%:0%:70%
10%:30%:0%:60%
…
80%:10%:0%:10%</td><td>0%:10%:10%:80%
0%:10%:20%:70%
0%:10%:30%:60%
…
0%:80%:10%:10%</td></tr>
<tr><td colspan="4">Quaternary blends (84)</td></tr>
<tr><td>10%:10%:10%:70%</td><td>10%:40%:40%:10%</td><td>20%:30%:40%:10%</td><td>30%:50%:10%:10%</td></tr>
<tr><td>10%:10%:20%:60%</td><td>10%:50%:10%:30%</td><td>20%:40%:10%:30%</td><td>40%:10%:10%:40%</td></tr>
<tr><td>10%:10%:30%:50%</td><td>10%:50%:20%:20%</td><td>20%:40%:20%:20%</td><td>40%:10%:20%:30%</td></tr>
<tr><td>10%:10%:40%:40%</td><td>10%:50%:30%:10%</td><td>20%:40%:30%:10%</td><td>40%:10%:30%:20%</td></tr>
<tr><td>10%:10%:50%:30%</td><td>10%:60%:10%:20%</td><td>20%:50%:10%:20%</td><td>40%:10%:40%:10%</td></tr>
<tr><td>10%:10%:60%:20%</td><td>10%:60%:20%:10%</td><td>20%:50%:20%:10%</td><td>40%:20%:10%:30%</td></tr>
<tr><td>10%:10%:70%:10%</td><td>10%:70%:10%:10%</td><td>20%:60%:10%:10%</td><td>40%:20%:20%:20%</td></tr>
<tr><td>10%:20%:10%:60%</td><td>20%:10%:10%:60%</td><td>30%:10%:10%:50%</td><td>40%:20%:30%:10%</td></tr>
</table>

Table S2-2-*continued*. Blended oil samples prepared for establishment of PLS-R models for quaternary blended oils

	Training set (286)				**Testing set (13)**
	Quaternary blends (84)				
Quaternary blended oils	10%:20%:20%:50%	20%:10%:20%:50%	30%:10%:20%:40%	40%:30%:10%:20%	
	10%:20%:30%:40%	20%:10%:30%:40%	30%:10%:30%:30%	40%:30%:20%:10%	
	10%:20%:40%:30%	20%:10%:40%:30%	30%:10%:40%:20%	40%:40%:10%:10%	
	10%:20%:50%:20%	20%:10%:50%:20%	30%:10%:50%:10%	50%:10%:10%:30%	
	10%:20%:60%:10%	20%:10%:60%:10%	30%:20%:10%:40%	50%:10%:20%:20%	
	10%:30%:10%:50%	20%:20%:10%:50%	30%:20%:20%:30%	50%:10%:30%:10%	
	10%:30%:20%:40%	20%:20%:20%:40%	30%:20%:30%:20%	50%:20%:10%:20%	
	10%:30%:30%:30%	20%:20%:30%:30%	30%:20%:40%:10%	50%:20%:20%:10%	
	10%:30%:40%:20%	20%:20%:40%:20%	30%:30%:10%:30%	50%:30%:10%:10%	
	10%:30%:50%:10%	20%:20%:50%:10%	30%:30%:20%:20%	60%:10%:10%:20%	
	10%:40%:10%:40%	20%:30%:10%:40%	30%:30%:30%:10%	60%:10%:20%:10%	
	10%:40%:20%:30%	20%:30%:20%:30%	30%:40%:10%:20%	60%:20%:10%:10%	
	10%:40%:30%:20%	20%:30%:30%:20%	30%:40%:20%:10%	70%:10%:10%:10%	

Table S3. PLS-R models of olive oil (OO) – sunflower seed oil (SF) blends determined by different criteria

Criterion	A	R^2Y	Q^2	Oil species	Training set			Testing set	
					RMSEE	RMSEcv	R^2	RMSEP	R^2
R(0.90)	1	0.9995	0.9995	OO	0.0075	0.0076	0.9995	0.0064	0.9996
R(0.95)	2	0.9998	0.9997	OO	0.0052	0.0056	0.9998	0.0058	0.9996
RMSE	3	0.9998	0.9997	OO	0.0049	0.0055	0.9998	0.0054	0.9997

*For information of the parameters, please see Section 2.5 in the manuscript.

Table S4. PLS-R models of olive oil (OO) – perilla oil (PR) – sunflower seed oil (SF) blends determined by different criteria

Criterion	A	R^2Y	Q^2	Oil species	Training set			Testing set	
					RMSEE	RMSEcv	R^2	RMSEP	R^2
R(0.90) &R(0.95)	5	0.9948	0.9946	OO	0.0181	0.0195	0.9956	0.0283	0.9953
				PR	0.0199	0.0211	0.9946	0.0357	0.9917
				SF	0.0206	0.0221	0.9943	0.0296	0.9951
RMSE	8	0.9956	0.9949	OO	0.0177	0.0195	0.9957	0.0282	0.9954
				PR	0.0180	0.0205	0.9956	0.0358	0.9917
				SF	0.0186	0.0212	0.9954	0.0298	0.9953

Table S5. PLS-R models of olive oil (OO) – perilla oil (PR) – rice bran oil (RB) – sunflower seed oil (SF) blends determined by different criteria

Criterion	A	R^2Y	Q^2	Oil species	Training set			Tesing set	
					RMSEE	RMSEcv	R^2	RMSEP	R^2
R(0.90)	8	0.9920	0.9918	OO	0.0181	0.0185	0.9938	0.0235	0.9939
				PR	0.0171	0.0173	0.9944	0.0239	0.9947
				RB	0.0262	0.0265	0.9868	0.0372	0.9900
				SF	0.0188	0.0192	0.9932	0.0331	0.9904
R(0.95)	10	0.9929	0.9927	OO	0.0170	0.0173	0.9945	0.0210	0.9956
				PR	0.0170	0.0171	0.9945	0.0221	0.9948
				RB	0.0238	0.0243	0.9891	0.0287	0.9929
				SF	0.0183	0.0187	0.9936	0.0320	0.9924
RMSE	15	0.9939	0.9935	OO	0.0158	0.0166	0.9953	0.0213	0.9961
				PR	0.0160	0.0167	0.9951	0.0224	0.9946
				RB	0.0219	0.0230	0.9908	0.0278	0.9926
				SF	0.0170	0.0176	0.9945	0.0315	0.9936

Table S6. Quantitative results of olive oil – sunflower seed oil blends measured on the same day and measured in 8 different days in 20 days

Actual olive oil con. (%)	Intra-day (n = 8)			Inter-day (n = 8)		
	Measured olive oil con. (%)	Relative error (%)	RSD (%)	Measured olive oil con. (%)	Relative error (%)	RSD (%)
7.8	7.7±0.4	–0.4	4.9	6.5±0.8	–16.1	11.8
29.9	29.9±0.5	–0.1	1.8	27.8±0.6	–7.0	2.0
49.8	49.6±0.6	–0.5	1.3	48.3±0.9	–3.2	1.9
69.7	69.6±0.9	–0.2	1.2	69.6±1.4	–0.1	2.0
91.6	91.2±0.4	–0.4	0.4	91.0±1.5	–0.6	1.6

Table S7. PLS-R model of olive oil (OO) – sunflower seed oil (SF) blends for inter-day measurements

A	R^2Y	Q^2	Oil species	Training set			Testing set	
				RMSEE	RMSEcv	R^2	RMSEP	R^2
3	0.9986	0.9975	OO	0.0135	0.0173	0.9986	0.0157	0.9973

Table S8. Quantitative results of olive oil (OO) – perilla oil (PR) – rice bran oil (RB) – sunflower seed oil (SF) blends by the full range and zoom-in models

Sample	Oil species	Actual con. (%)	Full range model			Zoom-in model		
			Measured con. (%)	Relative error (%)	RSD (%)	Measured con. (%)	Relative error (%)	RSD (%)
Q1	OO	100.0	95.2±0.9	–4.8	1.0	99.9±0.8	–0.1	0.8
	PR	0.0	1.5±0.8	/	/	0.0±0.6	/	/
	RB	0.0	0.0±1.4	/	/	–0.1±0.7	/	/
	SF	0.0	3.3±0.8	/	/	0.4±0.2	/	/
Q2	OO	0.0	3.5±1.7	/	/	0.0±0.3	/	/
	PR	100.0	94.6±0.6	–5.4	0.6	99.6±0.8	–0.4	0.8
	RB	0.0	2.7±3.0	/	/	0.9±1.1	/	/
	SF	0.0	–0.4±0.9	/	/	–0.5±0.9	/	/
Q3	OO	0.0	–2.4±1.1	/	/	0.3±0.4	/	/
	PR	0.0	–0.4±0.2	/	/	0.1±0.3	/	/
	RB	100.0	105±1.7	5.0	1.6	99.3±0.8	–0.7	0.8
	SF	0.0	–2.2±1.3	/	/	0.3±0.7	/	/
Q4	OO	0.0	–0.2±0.5	/	/	–0.5±0.4	/	/
	PR	0.0	3.7±0.2	/	/	0.2±0.2	/	/
	RB	0.0	5.8±1.8	/	/	1.6±0.7	/	/
	SF	0.0	90.6±1.5	–9.4	1.6	98.7±0.7	–1.3	0.7
Q5	OO	10.2	9.5±0.3	–7.5	2.9	9.4±0.5	–8.1	5.0
	PR	9.9	9.5±0.1	–4.0	0.8	9.1±0.2	–8.5	2.0
	RB	10.2	11.1±0.8	9.0	7.4	11.0±1.3	7.6	12.1
	SF	69.6	70.0±0.9	0.6	1.3	70.4±1.4	1.2	2.0
Q6	OO	10.1	8.9±0.4	–12.2	4.5	9.9±0.4	–2.0	4.4
	PR	10.6	8.5±0.5	–19.9	5.8	9.9±0.6	–7.2	6.5
	RB	69.4	72.2±1.0	4.1	1.3	69.9±0.9	0.8	1.3
	SF	9.9	10.2±0.7	3.1	6.4	10.3±0.8	4.3	7.5
Q7	OO	10.1	8.7±0.3	–13.9	3.3	9.3±0.4	–7.7	3.9
	PR	69.3	70.9±0.4	2.3	0.5	70.7±0.3	1.9	0.4
	RB	10.3	10.9±0.2	5.2	2.3	10.7±0.6	3.5	5.6
	SF	10.2	9.3±0.4	–9.3	4.2	9.3±0.5	–8.7	5.5
Q8	OO	70.1	70.9±0.5	1.3	0.6	72.7±0.3	3.8	0.5
	PR	10.1	8.9±0.2	–11.8	1.7	5.8±0.3	–42.5	5.4
	RB	9.9	10.1±0.5	2.9	5.3	10.2±0.5	3.7	4.8
	SF	10.0	10.1±0.3	1.0	3.1	11.3±0.3	12.9	2.6

Table S8-*continued*. Quantitative results of olive oil (OO) – perilla oil (PR) – rice bran oil (RB) – sunflower seed oil (SF) blends by the full range and zoom-in models

Sample	Oil species	Actual con. (%)	Full range model			Zoom-in model		
			Measured con. (%)	Relative error (%)	RSD (%)	Measured con. (%)	Relative error (%)	RSD (%)
Q9	OO	19.7	19.8±0.7	0.5	3.5	19.3±0.5	–2.0	2.4
	PR	20.4	19.8±0.3	–3.2	1.4	19.9±0.5	–2.5	2.3
	RB	20.2	22.3±1.0	10.4	4.7	25.5±1.2	26.1	4.7
	SF	39.6	38.3±0.8	–3.4	2.0	35.3±1.1	–11.0	3.1
Q10	OO	20.1	20.7±0.6	2.7	3.1	20.9±1.0	3.6	4.6
	PR	19.8	19.0±0.6	–4.1	3.0	19.9±0.6	0.2	3.2
	RB	40.3	39.5±1.0	–2.1	2.4	42.1±0.9	4.4	2.3
	SF	19.8	21.6±0.7	9.1	3.1	17.2±1.3	–12.9	7.4
Q11	OO	19.9	19.2±0.7	–3.7	3.4	17.2±0.3	–13.4	1.9
	PR	40.3	41.8±0.7	3.8	1.7	41.9±0.6	4.0	1.4
	RB	20.0	19.1±1.8	–4.5	9.2	19.7±1.4	–1.6	6.8
	SF	19.8	19.9±0.8	0.6	4.3	21.4±0.9	7.9	4.4
Q12	OO	39.9	40.7±0.6	2.0	1.5	43.6±0.9	9.2	2.1
	PR	19.8	18.6±0.6	–6.0	3.2	18.1±0.4	–8.6	2.2
	RB	20.2	19.1±1.0	–5.5	5.0	15.6±1.1	–22.4	7.1
	SF	20.1	21.6±0.6	7.4	2.7	22.6±0.6	12.6	2.7
Q13	OO	24.6	24.7±0.5	0.7	2.1	24.6±0.7	0.2	2.9
	PR	25.1	24.4±1.0	–2.7	3.9	25.0±0.7	–0.4	2.9
	RB	25.3	23.9±1.5	–5.8	6.4	26.0±1.2	2.6	4.7
	SF	25.0	27.0±0.8	7.9	3.1	24.4±0.5	–2.2	2.0

Table S9. PLS-R models of blended oils based on GC-FID chromatograms and MALDI-MS spectra

Type[a]	Method	A	R^2Y	Q^2	Oil species	Training set			Testing set	
						RMSEE	RMSEcv	R^2	RMSEP	R^2
OO-SF	GC	4	0.9994	0.9985	OO	0.0095	0.0134	0.9994	0.0297	0.9993
	MALDI	3	0.9995	0.9993	OO	0.0078	0.0095	0.9995	0.0093	0.9991
SF-CA-GP	GC	3	0.9652	0.9638	SF	0.0622	0.0629	0.9472	0.0888	0.9669
					CA	0.0244	0.0250	0.9916	0.0272	0.9938
					GP	0.0555	0.0555	0.9567	0.0957	0.9572
	MALDI	6	0.9832	0.9825	SF	0.0444	0.0450	0.9732	0.0524	0.9832
					CA	0.0212	0.0215	0.9938	0.0228	0.9967
					GP	0.0359	0.0362	0.9825	0.0381	0.9900

[a] OO: olive oil, SF: sunflower seed oil, CA: canola oil, GP: grapeseed oil.

Table S10. Fatty acid compositions of sunflower seed oil, canola oil and grapeseed oil as determined by GC-FID

Fatty acid	Sunflower seed oil	Canola oil	Grapeseed oil
C16:0	6.4%	4.4%	7.0%
C16:1	0.1%	0.2%	0.1%
C18:0	3.3%	1.9%	3.9%
C18:1	29.8%	61.9%	20.5%
C18:2	59.3%	20.7%	68.0%
C18:3, n6	0.0%	0.3%	0.0%
C18:3, n3	0.1%	8.3%	0.3%
C20:0	0.2%	0.6%	0.2%
C20:1	0.1%	1.3%	0.2%
C22:0	0.7%	0.3%	0.0%
C22:1	0.0%	0.2%	0.0%

Table S11. PLS-R models for determination of oil compositions of the commercial blended oil products

Model	Type[a]	A	R^2Y	Q^2	Oil species	Training set			Testing set	
						RMSEE	RMSEcv	R^2	RMSEP	R^2
M1	OO-SF	2	0.9972	0.9966	OO	0.0185	0.0197	0.9972	0.0167	0.9968
M2	OO-SF	2	0.9979	0.9975	OO	0.0161	0.0169	0.9979	0.0139	0.9990
M3	OO-CA	4	0.9985	0.9947	OO	0.0140	0.0261	0.9985	0.0170	0.9969
M4	OO-CO	3	0.9981	0.9971	OO	0.0156	0.0180	0.9981	0.0108	0.9987
M5	OO-CO	3	0.9984	0.9979	OO	0.0144	0.0149	0.9984	0.0107	0.9988
M6	OO-PA	4	0.9967	0.9945	OO	0.0205	0.0249	0.9967	0.0294	0.9899
M7	OO-PR	4	0.9973	0.9953	OO	0.0187	0.0245	0.9973	0.0152	0.9974
M8	OO-FS-SF	5	0.9985	0.9933	OO	0.0170	0.0188	0.9961	0.0202	0.9970
					FS	0.0251	0.0266	0.9914	0.0265	0.9966
					SF	0.0224	0.0243	0.9931	0.0262	0.9955
M9	OO-SO-SF	5	0.9900	0.9896	OO	0.0131	0.0133	0.9977	0.0146	0.9986
					SO	0.0317	0.0320	0.9861	0.0589	0.9844
					SF	0.0315	0.0321	0.9863	0.0578	0.9761

[a] OO: olive oil, SF: sunflower seed oil, CA: canola oil, CO: corn oil, PA: high oleic acid peanut oil, PR: perilla oil, FS: flaxseed oil, SO: soybean oil.

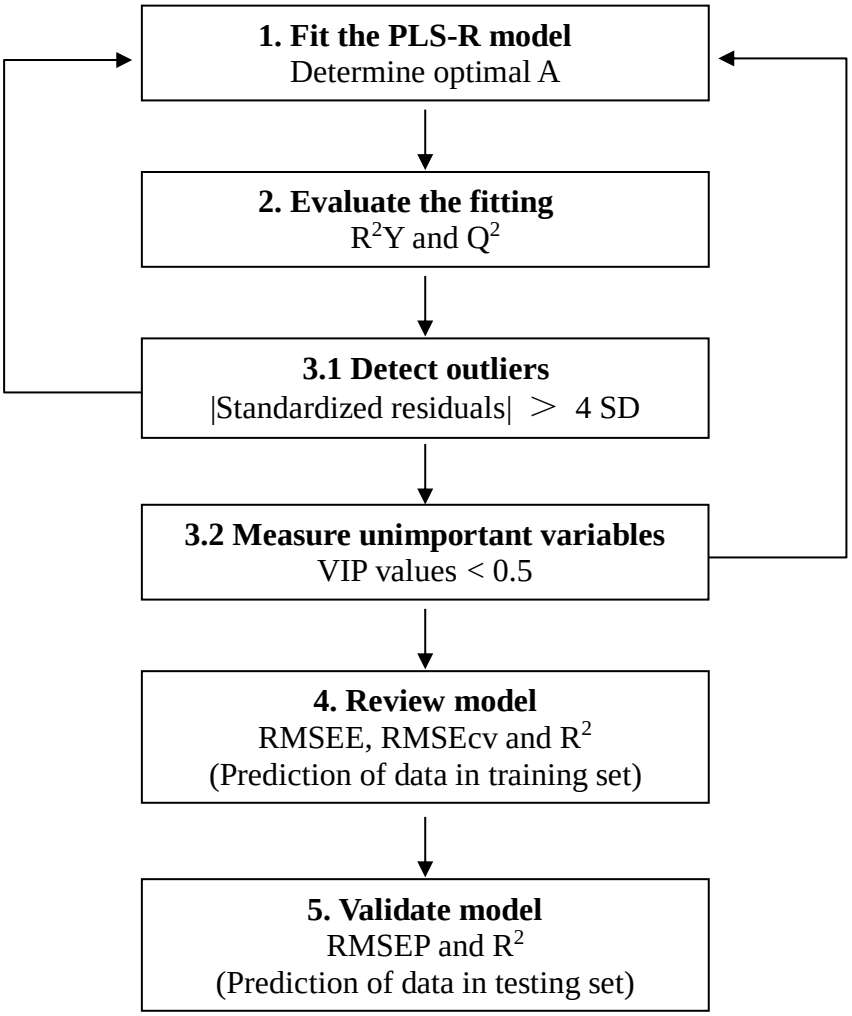

Figure S1. Protocol for establishing and optimizing the PLS-R models. For information of the parameters, please refer to Section 2.5 in the manuscript.

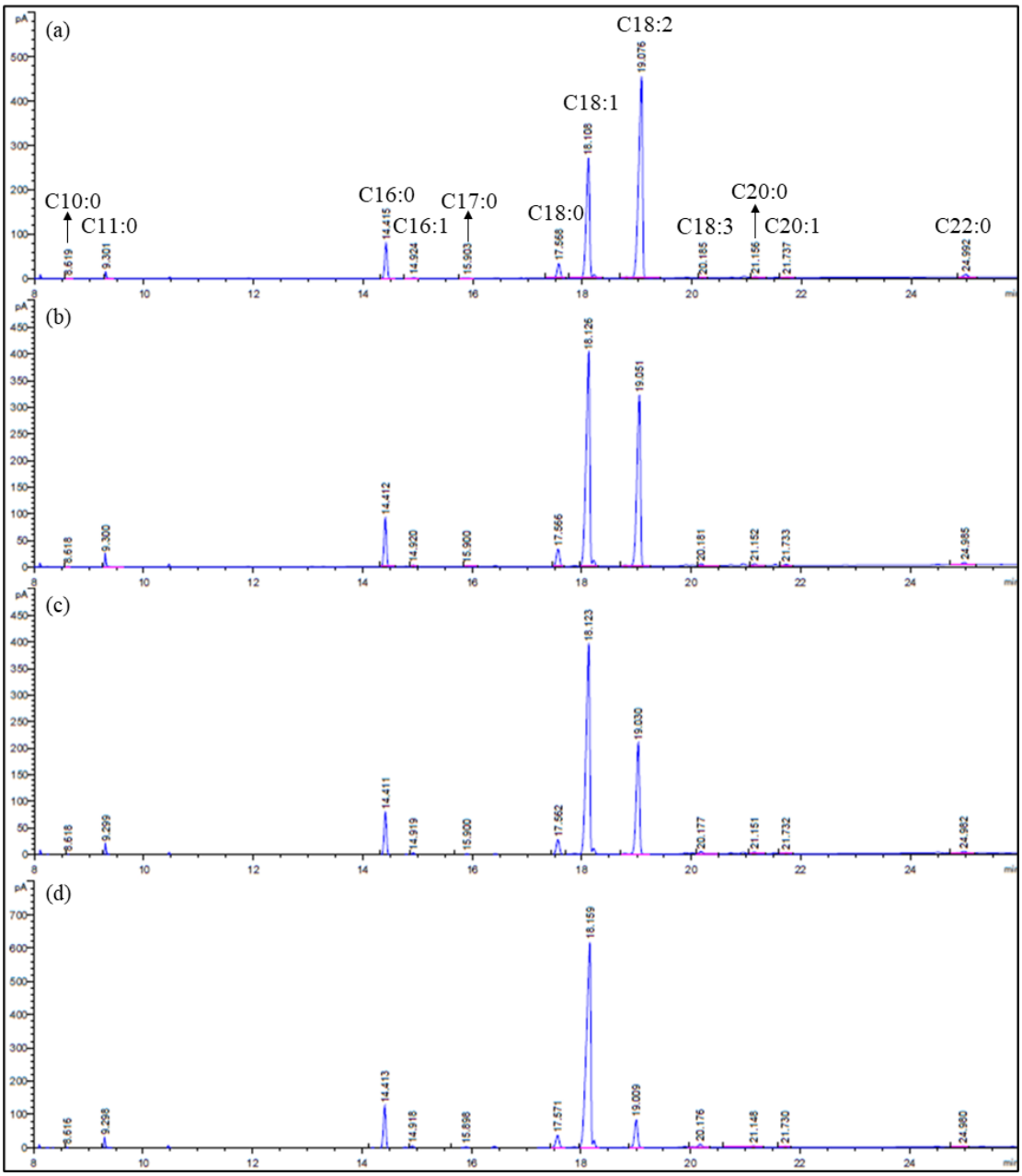

Figure S2. GC-FID chromatograms of (a) 100% sunflower seed oil, (b) 40% olive oil – 60% sunflower seed oil blend, (c) 60% olive oil – 40% sunflower seed oil blend, and (d) 100% olive oil.

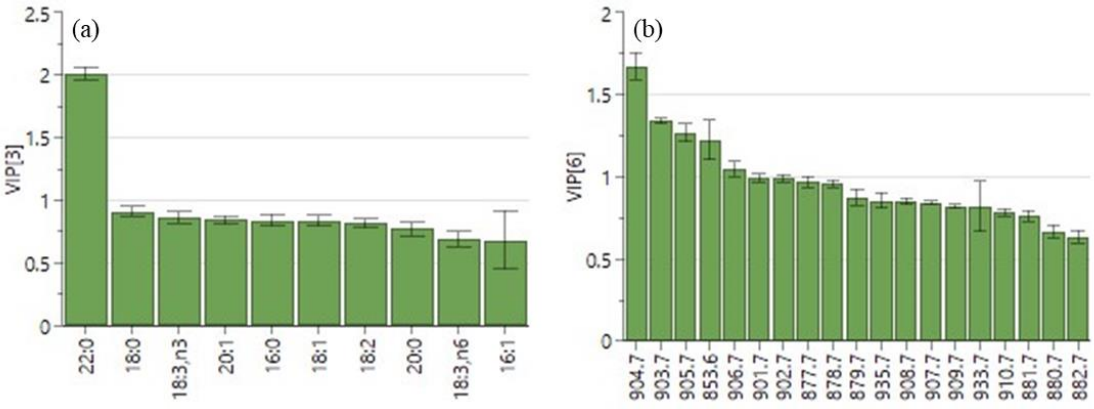

Figure S3. VIP values of PLS-R models based on (a) the GC-FID chromatograms and (b) the MALDI-MS spectra of sunflower seed oil – canola oil – grapeseed oil blends.

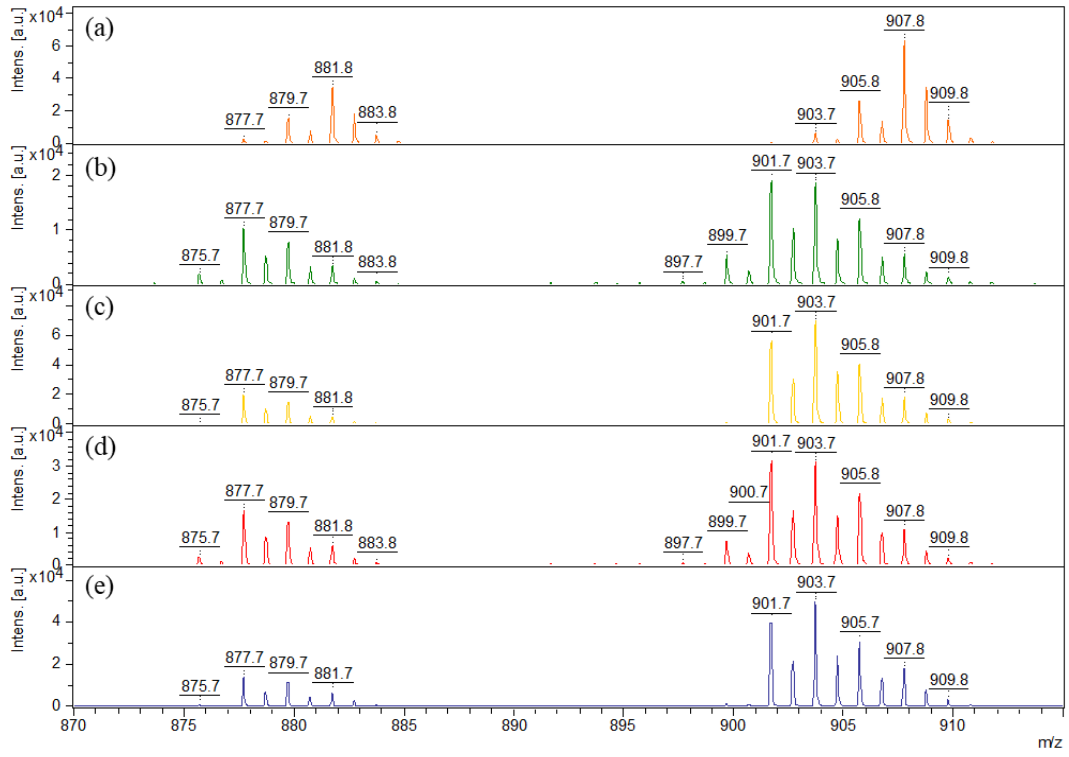

Figure S4. The TAG region of the MALDI-MS spectra for (a) 100% olive oil, (b) 100% soybean oil, (c) 100% sunflower seed oil, (d) commercial product 10, and (e) 10% olive oil – 10% soybean oil – 80% sunflower seed oil blend.