Anal Bioanal Chem (2011) 399:3491–3502
DOI 10.1007/s00216-011-4728-6

PAPER IN FOREFRONT

MALDI-MS and multivariate analysis for the detection and quantification of different milk species

Nicoletta Nicolaou · Yun Xu · Royston Goodacre

Received: 29 November 2010 / Revised: 25 January 2011 / Accepted: 25 January 2011 / Published online: 6 February 2011

Abstract The extensive consumption of milk and dairy products makes these foodstuffs targets for potential adulteration with financial gains for unscrupulous producers. Such practices must be detected as these can impact negatively on product quality, labelling and even health. Matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI-ToF-MS) is a potentially useful technique, with proven abilities in protein identification and more recently through the use of internal standards for quantification purposes of specific proteins or peptides. In the current work, we therefore aim to explore the accuracy and attributes of MALDI-ToF-MS with chemometrics for the detection and quantification of milk adulteration. Three binary mixtures containing cows' and goats', cows' and sheep's, and goats' and sheep's milk and a fourth tertiary mixture containing all types of milk were prepared and analysed directly using MALDI-ToF-MS. In these mixtures, the milk concentrations of each milk varied from 0% to 100% in 5% steps. Multivariate statistical methods including partial least squares (PLS) regression and non-linear Kernel PLS regression were employed for multivariate calibration and final interpretation of the results. The results for PLS and KPLS were encouraging with between 2% and 13% root mean squared error of prediction on independent data; KPLS slightly outperformed PLS. We believe that these results show that MALDI-ToF-MS has excellent potential for future use in the dairy industry as a rapid method of detection and enumeration in milk adulteration.

Keywords Bioanalytical methods · Chemometrics/statistics · Foods/beverages · Mass spectrometry/MALDI-MS

Electronic supplementary material The online version of this article (doi:10.1007/s00216-011-4728-6) contains supplementary material, which is available to authorized users.

N. Nicolaou · Y. Xu · R. Goodacre (✉)
School of Chemistry and Manchester Interdisciplinary Biocentre,
The University of Manchester,
131 Princess Street,
Manchester M1 7DN, UK
e-mail: roy.goodacre@manchester.ac.uk

Introduction

In today's era of consumerism and increasing reliability on the food industry to provide food fit for consumption, the issue of food safety and authenticity is becoming progressively more important. The internationalisation of food markets has also made the food industry a very competitive and financially lucrative business. The simple substitution of a scarce ingredient with a more abundant and cheaper ingredient, especially when high value products are involved, can yield huge financial gains. This however may also result in an unwanted and catastrophic chain of events as product quality usually becomes substandard, product identity is lost, incorrect product labelling ensues and consumers are exploited and mislead. At the same time, loss of the original high quality and origin-specific product may occur. Furthermore, incorrect product labelling may leave consumers exposed to potential allergens, such as casein protein from cows' milk, becoming detrimental for health.

Milk and dairy products are extensively consumed by large segments of the population during all stages of development and life, including childhood, adolescence, pregnancy and the elderly, due to their high nutritional value and health benefits. In general, however, milk is considered to be an expensive raw material. The current commercial UK price (from national outlets in 2010) for

cows' milk is ~£0.70/L, with sheep's over £3.00/L and goats' milk between £1.00/L and £1.34/L, with some seasonal milk production variation. This makes milk and dairy product adulteration, especially of the higher value types of milk, very profitable.

The variation of environmental conditions and the process of natural selection have also lead to the diversity of animal breeds and the production of milk with particular characteristics for that animal type and area [1]. Such milk has subsequently been used by local producers for the production of distinct types of cheeses of recognised quality and characteristics [2]. The production of these products however entails high overall processing costs, and producers can be financially destroyed by the presence of unfair competition, thus there may be a temptation to stretch the more expensive milk with one of lower value.

In an attempt to protect consumers and genuine product producers, policymakers at different countries have developed a number of legislations. The agriculturally diverse European Union has also legislated regulation EC/178/2002, a 'food safety and traceability regulation' [3], aiming to protect human life and health, establish consumer rights to food safety and accurate information and protect name misuse and imitation. European regulations on industrial milk processing are strict, only permitting a predefined number of constituent modifications, such as changes in the fat content and the addition of certain minerals, vitamins and milk proteins. Currently, certain dairy products, such as 'protected destination of origin' (PDO) and 'protected geographical indication' (PGI) goods, also require very accurate labelling in regards to their origin and processing and are protected by appellations of origin [4–6]. Further food labelling legislation is currently under discussion in the EU.

Ensuring product authenticity and implementing some of these strict standards and criteria has been very difficult. Analytical techniques have been employed to perform this hard task but have been unable to fullfill this role effectively, either because they are unable to keep in pace with the constant technological advances and developments concurring at the dairy industry or because of lack of commercial practicality [6]. Advanced techniques such as spectroscopy, near-infrared (NIR), mid-infrared (MIR) and nuclear magnetic resonance (NMR) spectroscopies [7–12], as well as chromatography [13, 14], immunoenzymatic assays [15, 16], polymerase chain reaction [17, 18], electrophoresis [19–21] and sensory analyses [22, 23], have all been utilised by the analytical dairy science to improve milk product analysis. The main disadvantage of these techniques though is that they remain time-consuming and labour intensive, with limited value for routine use in the screening of milk in the dairy industry.

Matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI-ToF-MS) is a potentially useful technique in the authentication of milk, with proven abilities in protein identification and more recently quantification [24–30]. MALDI is an ionisation technique involving the insertion of the sample typically into a UV absorbing matrix composed of a non-volatile material (usually a mild aromatic acid), followed by laser irradiation (typically at 337 nm), absorption, matrix energy desorption, and matrix to sample proton transfer resulting in the creation of vaporised ions. The technique has the advantage of only requiring small sample quantities and can be used for the analysis of heterogeneous biological samples such as milk, as recently demonstrated by Liland and colleagues [31, 32]. In addition, it possesses a very high sensitivity of mass range of up to 300,000 Da for proteins [33]. The aim of our study was therefore to investigate whether the MALDI-ToF-MS technique is able to detect and quantify goats' and sheep's milk adulteration with cows' milk, using the whole spectrum of peaks obtained from the analysis.

Materials and methods

Sample preparation

The fresh full fat pasteurised milk used in this study was purchased from national retail outlets and was analysed immediately. The milk tested included three types of milk: cows', goats' and sheep's milk. Four different milk type combinations were prepared as follows:

(1) Sheep's milk adulterated with cows' milk,
(2) Goats' milk adulterated with cows' milk,
(3) Sheep's milk adulterated with goats' milk,
(4) A tertiary mixture containing sheep's, goats' and cows' milk.

Each milk combination included a variation in the percentage of the primary milk adulteration ranging from 0% to 100% in successive increasing steps of 5%. A similar principle was used for the tertiary mixture (concentration levels are shown in Table S1 in the Electronic Supplementary Material). In order to ensure adequate mixing of the different milk types, the different mixtures were placed into sterile flasks and stirred using a rotational incubator for 15 min. Following this, 1 mL milk samples were collected and used for MALDI-ToF-MS analysis.

MALDI-ToF-MS

The sample preparation used was based on the method reported by Cozzolino and colleagues [26]. This involved initially taking 100 μL of each homogenized sample and

diluting these in 1 mL of water containing 0.1% trifloroacetic acid (Sigma-Aldrich, Dorset, UK). These samples were then further diluted 1:10 with the same solvent. Five microlitres of these samples were then mixed with 5 μL of the matrix solution. For all samples, sinapinic acid (Sigma-Aldrich, Dorset, UK) was used as a matrix having been saturated in a matrix solution composed of 50% acetonitrile (Sigma-Aldrich, Dorset, UK) and 50% water. From the final mixtures, 1 μL was then positioned onto a MALDI-MS stainless steel target plate and dried for 2 h at room temperature.

MALDI-ToF-MS analysis was undertaken on a MALDI-ToF mass spectrometer (AXIMA-CRF™*plus*; Shimadzu Biotech, Manchester, UK), equipped with a nitrogen pulsed UV laser (337 nm), and was operated using a positive ion source in linear ion mode in the positive ion mode. The laser power was set to 120 mV, each spot was analysed using a random raster of 500 profiles and each profile contained data from five laser shots. Each sample was analysed three times, and the typical collection times were 4 min per sample.

Data analysis

Pre-processing

The mass spectral data were imported into MATLAB (The Math Works, Natick, MA, USA) and processed for analysis. Typically, the data were baseline corrected and normalised. Normalization of each individual spectrum was performed by dividing each individual baseline corrected spectrum with the square root of the sum of squares of the spectrum [34].

Exploratory analysis

The exploratory analysis was performed in two steps as described by us elsewhere [12, 35]. Initially, principal component analysis (PCA) was used. PCA is an established procedure for reducing the dimensionality of multivariate MALDI-ToF-MS data whilst preserving most of the variance; this process results in the creation of new variables named principal components PCs which are uncorrelated. This is important because there are a large number of variables in the MALDI-ToF-MS data. The second step involved the use of canonical correlation analysis (CCA). CCA [36] is a commonly used method for assessing the correlation between two multivariate matrices or one multivariate matrix and one corresponding vector (e.g. concentration of adulterant). CCA seeks a set of linear combinations called canonical variables so that the correlation between the two matrices is maximised. The correlation of the two matrices is expressed as a correlation coefficient in a similar sense of the correlation coefficient (R) between two vectors, while the significance level of such correlation can be asses by using an F test [37]. CCA thus gives us a quick assessment of the correlation (R) between the MALDI-ToF-MS spectra and the adulteration level of the milk and the significance of that probability (p value) before we move to a more robust quantitative analysis. The CCA is performed on the scores from the PCA, and the results shown are based on the number of PCs which yielded the lowest p value.

Quantitative analysis

If there is a strong correlation between the two inputs, i.e. the MALDI-ToF-MS spectra and the adulteration levels, it is then possible to employ a multivariate regression model to predict the adulteration levels using the MALDI-ToF-MS spectra. In this study, we used partial least squares (PLS) and Kernel PLS, as linear and non-linear regression techniques, respectively.

Whilst supervised methods are very powerful, it is possible to over-fit the model; therefore, validation of both PLS and Kernel PLS was undertaken. We achieved this using an independent test set for each of the mixture types. For each of the binary mixtures, the training set contained 0%, 10%, 20%,..., 90% and 100% of one of the milks, and the test set included 5%, 15%, 25%,..., 85% and 95%. For the tertiary mixtures, the training and test sets are shown in Table S1. The choice of the training set and the test set is to ensure that both training and test set have a similar coverage of the adulteration levels except that in the test set there was no extrapolation (i.e. there is no pure milk of any kind to be predicted in the test set).

PLS regression

PLS regression is a frequently used multivariate algorithm [38]. It is more useful than standard multivariate regression because PLS is able to deal with multicollinearity in data; this is usual when continuous data are generated (such as those mass spectra shown in Fig. 1). PLS models were generated to predict a single variable, and so, PLS1 was used for both the three binary mixtures as well as for the tertiary mixture. The number of PLS components (latent variables) was optimized using a k-fold cross-validation on the training set only, while k is the number of adulteration levels in the training set.

Kernel PLS

Kernel PLS is a non-linear extension of PLS regression which makes use of Kernel learning concepts [39]. In Kernel learning, one projects the data into an appropriate

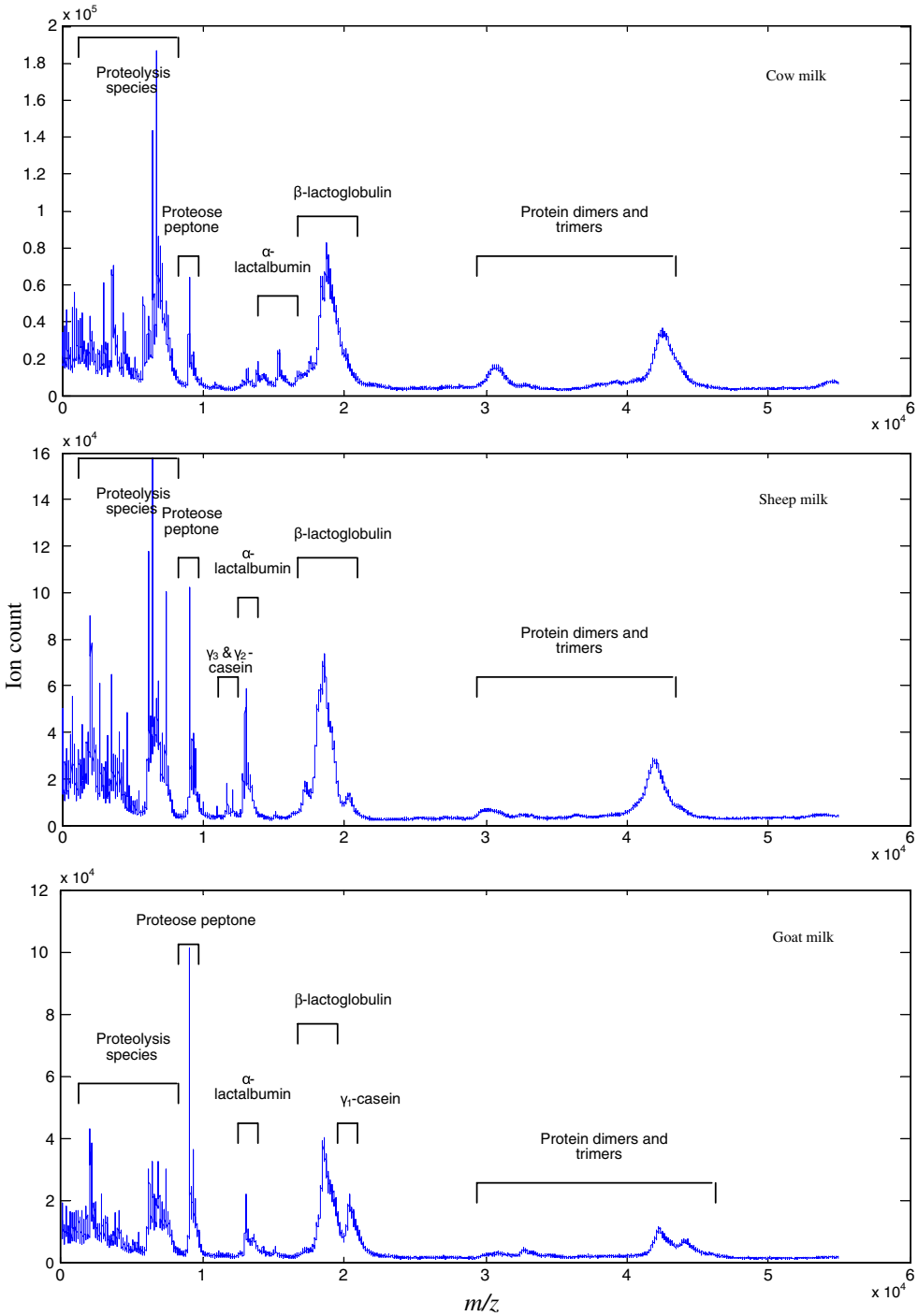

Fig. 1 Raw MALDI-ToF mass spectra of cows', sheep's and goats' milk samples. These mass spectra are annotated with protein identifications

higher dimensional feature space, with the result that many non-linear problems can now be solved by using a linear modelling method on this projected feature space [39]. In our KPLS approach, we used a radial base function (RBF) as the Kernel function, and the optimal combination of the Kernel parameter and the number of PLS factors were optimized using a grid search approach again using *k*-fold cross-validation on the training set as described above.

Results and discussion

Spectra

MALDI spectra can be used to determine and quantify the protein components of various types of milk, by identifying the different peaks and assigning them to specific proteins based on previously published protein molecular mass data. Individual proteins, however, may show a variation in their

molecular mass and thus the exact position of the peak; this is due to factors such as genetic and non-genetic polymorphisms and milk processing, the latter mainly via thermal denaturation and proteolysis which can affect individual protein structure [40–44]. Milk samples from different species or from different animal breeds of the same species can therefore display small variations in the molecular mass of the same protein, which will also vary depending on whether the milk is raw or has been processed and how it has been processed.

Figure S1 (in the Electronic Supplementary Material) shows a typical MALDI mass spectrum of fresh full fat pasteurised cow milk, including the original raw data and processed data after baseline correction and normalization that was required before chemometric analyses. Qualitatively, the spectra from all three milk species (Fig. 1) appear to display similar protein patterns between the different types of milk with small differences in the location of the molecular mass signal of the same proteins; in addition, it is also possible to see some differences in regards to the quantity of certain proteins. A closer inspection of the MALDI mass spectrum of cows' milk (Fig. 1) and interpretation based on previously published molecular mass data [24, 26, 27, 30, 45] reveal a number of protein-related peaks. These include a peak at ~*m/z*9,000 representing the proteose peptone, seen in all milk types at a similar position, and the peaks at ~*m/z* 15,000 and *m/z*18,500 relating to α-lactalbumin and β-lactoglobulin, respectively, with the latter displaying a higher peak/content compared to the other milk types. The broad peaks over *m/z*30,000, at *m/z*31,000 and *m/z*43,000 represent dimeric and trimeric species [45].

Inspection of the MALDI spectrum of sheep's milk shows the α-lactalbumin and β-lactoglobulin peaks appearing as ~*m/z*13,500 and ~*m/z*19,000, with an additional peak at ~*m/z*12,000 representing γ_2-casein, which is less prominent in the other milk types. Finally, in the MALDI spectrum of pure goats' milk, the α-lactalbumin and β-lactoglobulin are located at ~*m/z*13,000 and *m/z*19,000 peaks, respectively, while a more prominent peak at ~*m/z* 21,000 represents γ_1-casein.

The peaks below *m/z*9,000 in all types of milk are representative of species with low molecular mass due to proteolysis of higher mass proteins. The latter appears to be significant and over-represented in the MALDI spectra obtained in this study compared to spectra from other studies using raw milk, and this is most likely due to the effect of milk processing and pasteurisation on the various protein species.

Analysis of binary mixtures of milk

The small differences between the pure spectra became even harder to visualise by eye when mixtures were analysed, and so, this did not allow direct visual comparison in order to detect the level of adulteration. Therefore, the relationship between the spectra collected from the different milk combinations and concentrations was investigated using the exploratory analysis procedures as described in the "Data analysis" section.

The results from the CCA for mixtures of cows' and goats' milk analysed using MALDI-ToF-MS are shown in Fig. 2. It is very clear from this plot that there is a concentration-dependent relationship in the MALDI-ToF-MS data, and visual inspection of these results identifies a linear pattern of the mixture levels. The canonical correlation coefficient R for this linear relationship was found to be 0.9953 with a highly significant probability p value of 8.83×10^{-35}, while similar values (R=0.9893, p value=4.07×10^{-30}) were found for the cows' and sheep's milk mixtures (CCA plots not shown). The goats' and sheep's milk mixtures displayed a comparable linear relationship to the other milk mixtures but with a slightly lower value for the correlation coefficient, R=0.9674, and a p value of 2.70×10^{-16}. This suggests that there are very strong correlations between the MALDI-ToF-MS data and their corresponding milk adulteration levels. Linear and non-linear multivariate regression techniques were therefore employed in order to explore these trends even further and to assess whether it was possible to quantify the level of milk adulteration from these mass spectra.

Quantification of binary milk mixtures

Three different binary milk combinations were created as detailed above. The cow–goat milk binary mixture is

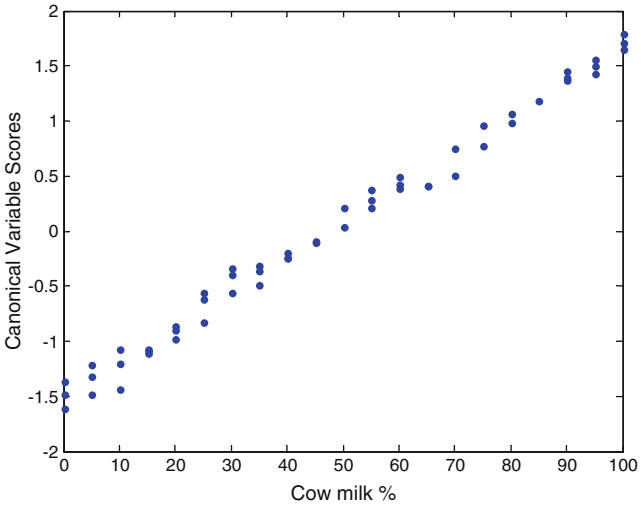

Fig. 2 CCA plot on MALDI-ToF mass spectra of cows' milk when added to goats' milk. PCs 1–15 were used by the CCA algorithm with a priori knowledge of concentration of cow milk. The *different dots* show the concentration of cows' milk in the mixture in relation to the canonical variable scores, with R=0.9953 and p value=8.83×10^{-35}

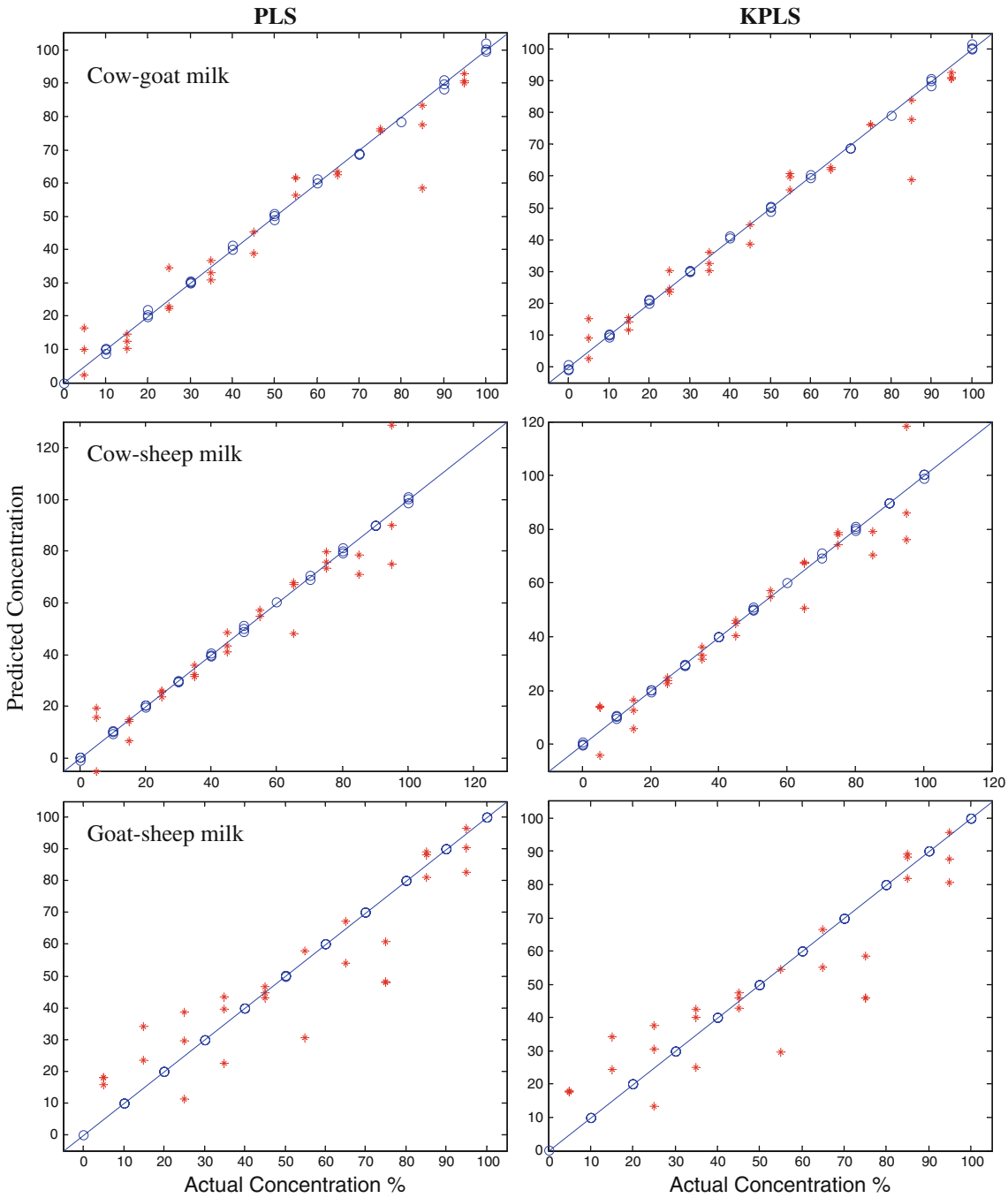

Fig. 3 Plots showing the predicted levels of milk adulteration estimated from PLS (*left column*) and KPLS (*right column*) models. The *rows* show the different mixtures analysed with predictions for cow's milk adulteration in goats' (*top*) and sheep's (*middle*) milk, and goats' milk adulterated into sheep's milk in the *bottom row*. The *blue circles* represent the training data set and the *red crosses*, the test set

subsequently used as an example. Samples containing 0% to 100% cows' milk (in 5% steps) in goats' milk underwent preparation, and the resulting 21 mixtures produced were analyzed in triplicate using MALDI-ToF-MS. As detailed above, the data were baseline corrected, normalised and then split into a training set (0%, 10%, 20%,…, 90% and 100% goats' milk) and a test set (5%, 15%, 25%,..., 85% and 95%) and analyzed by linear PLS and non-linear KPLS.

During calibration of the PLS model, sub-sampling of the training data took place, leaving one set (i.e. a unique adulteration level; all replicates) out so that a cross-validation set could later be generated; this allowed for the selection of the optimum number of PLS factors for calibration. Once this was performed, the independent test set was utilized to challenge the derived PLS model. PLS regression results for the cows' and goats' milk mixture

Table 1 Comparison of the partial least squares (PLS) regression and the non-linear Kernel partial least squares (KPLS) results of the MALDI-ToF mass spectra for determining the percentage volume of cows' milk mixed with goats' milk and sheep's milk and goats' milk mixed with sheep's milk

	PLS	Kernel PLS
Cow–goat milk		
Factors	7	7
RMSECV	5.24	4.98
RMSEC	0.95	0.77
RMSEP	6.85	6.35
Correlation coefficient in the train set (R^2)	0.99	0.99
Correlation coefficient in the test set (Q^2)	0.95	0.95
Cow–sheep milk		
Factors	9	9
RMSECV	7.56	6.16
RMSEC	0.64	0.51
RMSEP	9.77	8.13
Correlation coefficient in the train set (R^2)	0.99	0.99
Correlation coefficient in the test set (Q^2)	0.89	0.92
Goat–sheep milk		
Factors	18	19
RMSECV	10.59	10.35
RMSEC	0.02	0.03
RMSEP	12.38	12.87
Correlation coefficient in the train set (R^2)	1.0	1.0
Correlation coefficient in the test set (Q^2)	0.82	0.81

The RMSECV represents the root mean square error for the cross-validation, the RMSEC represents the root mean square error for the calibration and the RMSEP represents the root mean square error for the predictions produced for the independent test set

combinations are shown in Fig. 3. The plot of the estimated cows' milk concentration versus the known cows' milk concentration values in this figure appeared to show good predictive values, and importantly, both the training and test sets lie on the expected $y=x$ perfect prediction line. Table 1 shows the detailed results for all the three binary mixtures. For the cows'–goats' mixture, the root mean squared (RMS) error for the training data (RMSEC) was 0.95%, with seven PLS factors selected by the cross-validation set for this model, and the RMS error for the cross-validation (RMSECV) set was 5.24%; the RMS error in the independent test was 6.85% (RMSEP). This models' Q^2 value was 0.95, while the R^2 for the train set model was 0.99, both very close to the prefect model, which would be 1.

The most discriminative features used by the PLS regression, which models the adulteration levels of the milks, can be found by inspecting the variable importance for projection (VIP) plots. The most discriminative features generally have higher magnitudes than the non-discriminative ones in the VIP plots. These are displayed in Fig. 4 and Table 3 for the binary milk mixtures. Comparison with the pure milk MALDI-MS spectra (Fig. 1) indicates that for the cow–goat binary mixtures, the dominant features at 14,100 and 18,020 *m/z* are

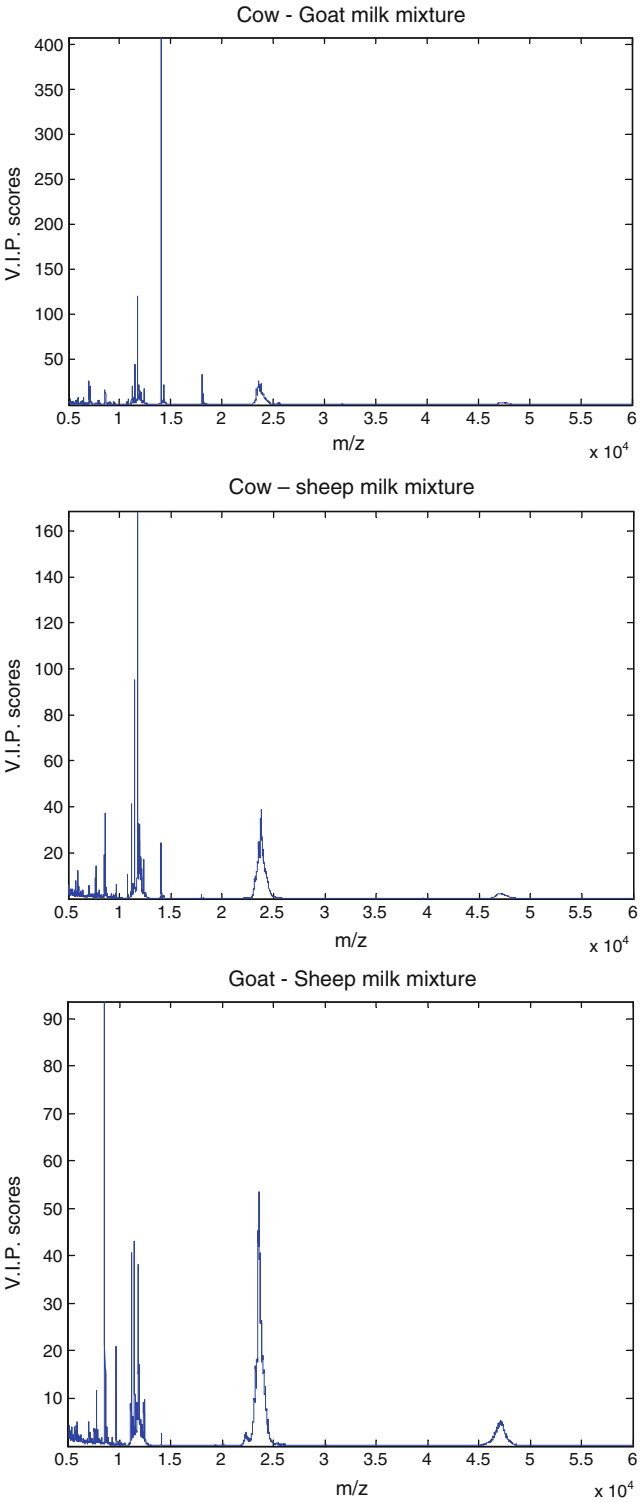

Fig. 4 Plots of variable importance for the prediction (VIP) scores used in the PLS modelling for the binary milk mixtures

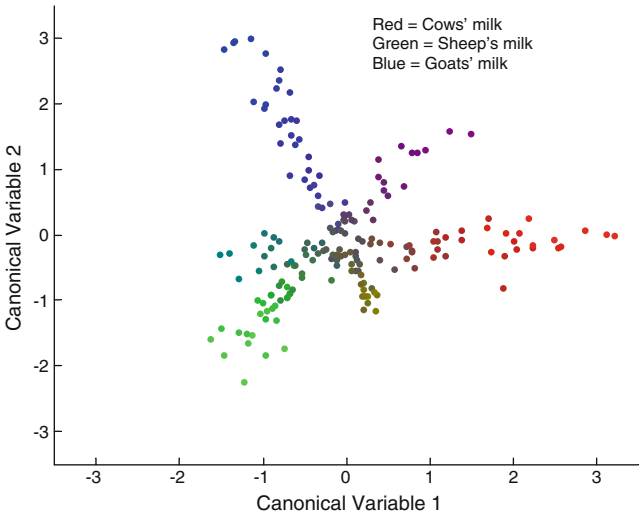

Fig. 5 CCA scores plot on MALDI-ToF mass spectra for the mixtures from the three types of milk. The first five PCs from the PCA were used for CCA with a priori knowledge of the concentration of the three milks. Pure cows' milk samples are presented with *red colour*, sheep's milk is presented with *green colour* and goat's milk is presented with *blue*. The colour of the mixtures of these three types of milk are represented by combining the three corresponding colours based on their relative concentration accordingly, e.g. a "*purple*" colour which is a mixture of red and blue with low green represents a mixture of goat and cow's milk with little to no sheep's milk. In the *middle*, there appears to be samples which contain two out of the three milk types in high concentrations, and these appear with different colours from the combination of all three types of milk

clearly present in both the pure cows' and goats' milk spectra, and these peaks represent α-lactalbumin and β-lactoglobulin, respectively. Whilst these proteins are present in both milks, the ratio of β-lactoglobulin to α-lactalbumin is different between cow and goat milk, and this is what is discriminatory. An additional peak at 11,740*m/z*, representing γ_2-casein, is also selected as a VIP, although it does not appear as a dominant peak in any of the two pure milk spectra; PLS suggest that despite its small magnitude, this protein is significantly different. In the cow–sheep binary mixture, VIP features at 8,600 and 14,100*m/z* appear in both the pure cows' and sheep's milk spectra representing the proteose peptone and α-lactalbumin proteins, respectively, while features at 11,190 and 11,280*m/z* (γ_3-casein) and 11,740*m/z* (γ_2-casein) are only present in the pure sheep's milk spectra. In the third binary mixture, goat–sheep milk, the latter features (γ_3-casein and γ_2-casein) remain distinctive only in the pure sheep's milk spectra with the VIP feature at ~8,600*m/z* (proteose peptone) being present in both of the pure milk spectra. The 23,470–700*m/z* feature (α_{S1}-casein), even though appearing as a dominant VIP feature in all three binary mixture PLS models, does not appear as an intense peak in the pure milk spectra for any of the three types of milk.

Considering the relationship between the response from the MALDI analysis and the corresponding adulteration levels of the milk may not be linear, we decided to employ the KPLS algorithm, a non-linear extension of PLS. The results are depicted in Fig. 3, with the relevant statistical information also shown in Table 1. For our cows'–goats' binary milk mixture, it is clear that KPLS provides similar results to PLS, with a RMSEP of 6.35% and a Q^2 of 0.95. The usefulness and improvement of the non-linear KPLS algorithm compared to PLS are more apparent when the results from the other two binary milk mixtures are observed, as shown in Fig. 3 and Table 1. KPLS does not allow the generation of loadings or VIP scores, so model interpretation in terms of which proteins are important is largely hidden.

Mixtures of the three types of milk

The MALDI mass spectra collected from the 63 milk samples, containing various concentrations of the three milk types, sheep's, goats' and cows' milk, were analyzed using a similar strategy to the binary mixture analysis. Initially, the relationship between the spectra and the adulteration levels of the milks was investigated using CCA. Figure 5 presents the CCA results in a pseudo-2D plot constructed from the first five PCs. The different concentrations of the three different types of milk are represented in different colours. We used red, green and blue for pure cows', sheep's and goats' milk, respectively. The colours of the mixture samples are represented by mixing the three colours together according to their corresponding relative concentration levels. For example, the colour a mixture sample of cows', sheep's and goats' milk with their relative concentration levels of 70%:20%:10% is a colour with its RGB (red, green and blue) channels having 7:2:1 relative intensities, respectively. Visual inspection of the 2D space indicates a clear pattern of distribution of the different milk mixture concentrations. Domination of a particular type of milk in the mixture at concentrations of greater than 50% appears to create a tentacle towards a specific direction with a pure milk type at the tip. An increasing cows' milk concentration appears to create a tentacle towards the right of the pane (red), an increase in sheep's milk concentration extends a tentacle towards the left lower corner of the pane (green), while similarly, the increasing goats' milk concentration forms a tentacle towards the upper left corner of the pane (blue). Furthermore, three additional smaller distinctive tentacles (yellow, purple and turquoise) are observed towards the middle of the pane extending outwards in different directions. Each tentacle appears to lie between two of the bigger tentacles and represents mixtures containing the two associated milk types in concentrations of greater than 40%, respectively. For example, the small tentacle in turquoise colour extending due west lies between the blue (high goat's milk concentration) and the

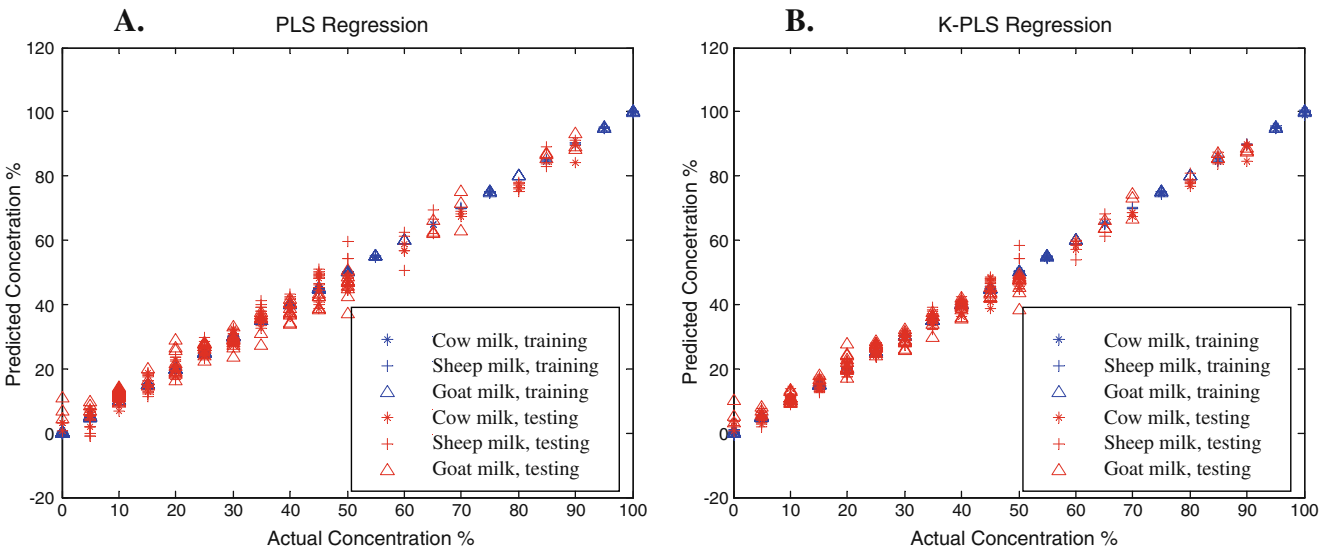

Fig. 6 Plots showing the predicted concentrations of cows', goats' and sheep's milk versus the actual concentrations for the training and test set in the tertiary mixtures of the three types of milk using (**a**) PLS and (**b**) KPLS. The *blue colour* represents the training data set and the *red*, the test set. The *triangles* represent the goats' milk samples, the *crosses*, the sheep's milk and the *stars*, the cows' milk

green tentacle (high sheep's milk concentration), representing mixtures containing these two types of milk in concentrations of greater than 40%. At the tip of the small tentacles lie the two dominant milk types at 50% concentration each. Even though this analysis of tertiary mixtures using CCA appears to show clear trends regarding the different milk combinations, because these are revealed in their 2D spatial distribution, this would limit accurate quantification, and thus PLS and KPLS were also used to quantify the levels of the different milk species in these tertiary mixtures.

Quantification of tertiary milk mixtures

In order to perform quantification of the tertiary mixtures, the data were first divided in two sets, a training and a test set (Table S1), and analyses were undertaken as described above again employing the linear and non-linear supervised

Table 2 Comparison of the partial least squares (PLS) regression and the non-linear Kernel partial least squares (KPLS) results of the MALDI-ToF mass spectra for determining the percentage volume of cows', goats' and sheep's milk from mixtures containing the three types of milk together

	Cow milk	Sheep milk	Goat milk
PLS			
Factors	9	10	9
RMSECV	2.58	4.77	4.25
RMSEC	0.19	0.12	0.10
RMSEP	2.42	3.29	3.84
Correlation coefficient in the train set (R^2)	0.99	0.99	0.99
Correlation coefficient in the test set (Q^2)	0.98	0.97	0.97
KPLS			
Factors	6	10	10
RMSECV	1.91	2.84	3.27
RMSEC	0.50	0.12	0.08
RMSEP	2.02	2.32	2.98
Correlation coefficient in the train set (R^2)	0.99	0.99	0.99
Correlation coefficient in the test set (Q^2)	0.99	0.98	0.98

The RMSECV represents the root mean square error for the cross-validation, the RMSEC represents the root mean square error for the calibration and the RMSEP represents the root mean square error for the predictions produced for the training set. The R^2 represents the correlation coefficient for the training set, and the Q^2 represents the test set

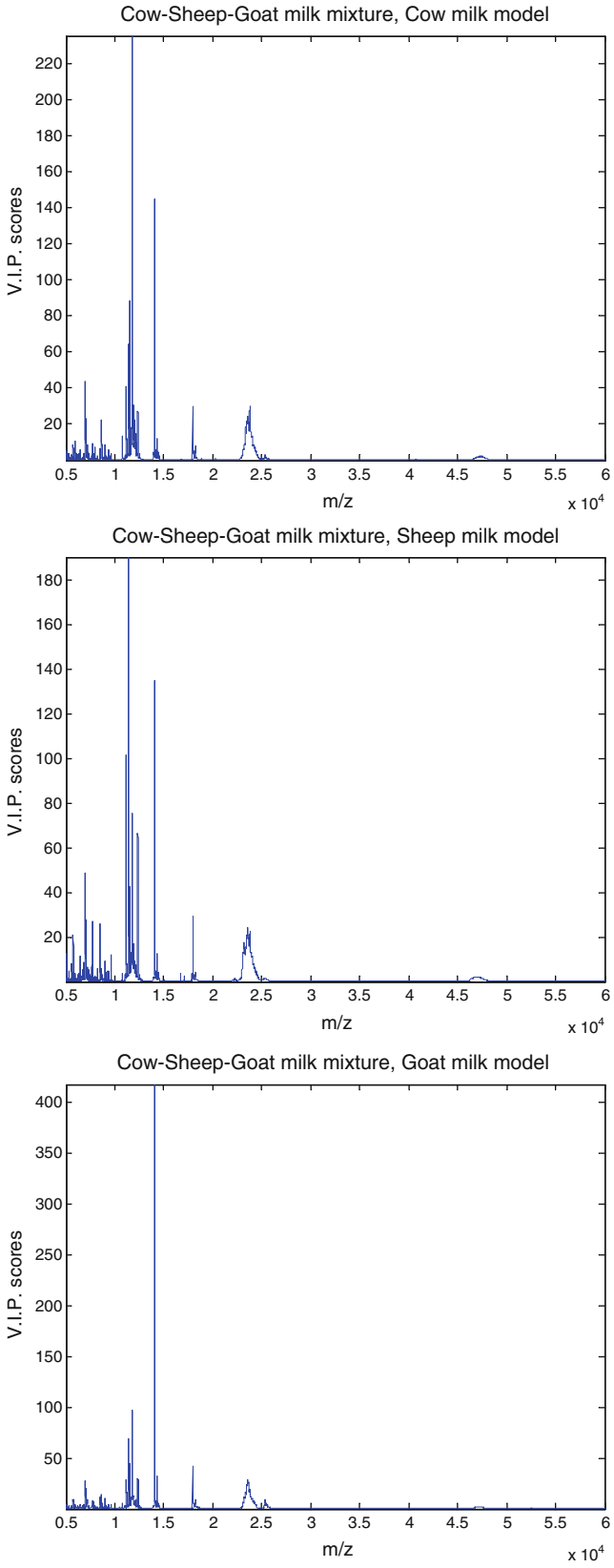

Fig. 7 Plots of variable importance for the prediction (VIP) scores used in the PLS modelling for the tertiary milk mixtures

Table 3 Spectral peaks with the highest variable importance for the prediction (VIP) scores used in the PLS modelling for the binary and tertiary milk mixtures

	Peak *m/z*	Protein
Binary mixture		
Cow–goat milk	11,750	γ_2-Casein
	14,100	α-Lactalbumin
	18,020	β-Lactoglobulin
	23,700	α_{S1}-Casein
Cow–sheep milk	8,600	Proteose peptone
	11,190, 11,480	γ_3-Casein
	11,740	γ_2-Casein
	23,470	α_{S1}-Casein
Goat–sheep milk	8,530	Proteose peptone
	11,190, 11,450	γ_3-Casein
	11,820	γ_2-Casein
	23,600	α_{S1}-Casein
Tertiary mixture		
Cow milk	11,470	γ_3-Casein
	11,730	γ_2-Casein
	14,060	α-Lactalbumin
Sheep milk	11,190, 11,450	γ_3-Casein
	14,060	α-Lactalbumin
Goat milk	11,450	γ_3-Casein
	11,740	γ_2-Casein
	14,060	α-Lactalbumin

learning techniques of PLS and KPLS regression. Although there are three Y-variables to be predicted (one for each milk species), rather than use PLS2 and KPLS2, we chose to use PLS1 and KPLS1 where three models were constructed for each milk. This was because it has been shown that PLS1 generally outperforms PLS2 for quantification of different analytes, as there will be different directions in the spectral space that are describing the contributions for the three different milk species; therefore, it is better to optimise each of these individually [46]. Figure 6 illustrates the results from the PLS and KPLS models for the training and test data. For both PLS and KPLS models, excellent predictions for the three different types of milk was attained. Table 2 shows the summary statistics for both multivariate regression models using the same training and test set splits; in this case, the PLS algorithm outperforms KPLS. Rather interestingly, these models for the tertiary mixture were better than the models constructed from the binary models (Fig. 3 and Table 1), although we cannot think of any conceivable reason why this may be the case.

The most dominant spectral peaks used by the PLS modelling for discriminating between the three types of milk with the highest VIP scores for the tertiary milk

mixtures are displayed in Fig. 7 and Table 3. In all three PLS models for each type of milk, the dominant features are largely the same and appear as a peak in one or more of the pure milk spectra. The features at 11,190–450 *m/z* (γ_3-casein) and 11,730–40 *m/z* (γ_2-casein) are only present as spectral peaks in the pure sheep milk, while the 14,060 *m/z* (α-lactalbumin) VIP score feature is present as a peak in all three types of pure milk samples. Furthermore, the difference found between the number and level of scoring for some features in the tertiary sample mixtures compared to the binary mixtures is most likely due to the effect of the presence of the additional milk type in the mixture.

Concluding remarks

In comparison to previous studies demonstrating the qualitative aspects of MALDI-ToF-MS using selected peaks for milk speciation [24–30], through this study, we have shown that the *whole* MALDI-ToF mass spectra contain valuable information. However, this can only be revealed when MS is combined with multivariate techniques such as linear PLS and non-linear Kernel PLS, and we have shown that it is possible to achieve very accurate predictions of the levels of milk species adulteration. These properties have been demonstrated in analysing binary and tertiary mixtures of fresh pasteurised cows', sheep's and goats' milk using a simple and fast process. Despite the milk processing, which may affects protein structure, MALDI-ToF-MS was able to achieve high accuracy levels of milk adulteration with typical errors in the region of 2–10% for cow's milk, a level at which a fraudster would unlikely adulterate at because this would not be financially viable.

Acknowledgments YX and RG acknowledge the Symbiosis-EU (http://www.symbiosis-eu.net) project (no. 211638) financed by the European Commission under the 7th Framework programme for RTD. The information in this document reflects only the authors' views and the Community is not liable for any use that may be made of the information contained therein.

References

1. Romero C, Perez-Andujar O, Olmedo A, Jimenez S (1996) Detection of cow's milk in ewe's or goat's milk by HPLC. Chromatographia 42(3/4):181–184
2. Summer A, Franceschi P, Bollini A, Formggioni P, Tosi F, Mariani P (2003) Seasonal variations of milk characteristics and cheese making losses in the manufacture of Parmigiano-Reggiano cheese. Vet Res Commun 27:663–666
3. EC EC (2002) General food law. Available at http://ec.europa.eu/food/food/foodlaw/index_en.htm. Accessed 2010
4. Bosset JO, Berger T, Buhler-Moor U, Collomb M, Dafflon O, Gauch R, Jeangros B, Lavanchy P, Mariaca R, Schovic J, Sieber R, Troxler J (1997) Comparison of some highland and lowland Gruyere type cheese of Switzerland: a study of their potential PDO/AOC/AOP characteristics. Authenticity and adulteration of food—the analytical approach. Proceedings of the European Food Chemistry IX 2 (FECS-Event No 220), pp 395–400
5. Bosset JO, Jeangros B, Berger T, Butikofer U, Collomb M, Gauch R, Lavanchy P, Scehovic J, Troxler J, Sieber R (1998) Comparison of Swiss hard cheese Gruyere-type produced and lowland. Rev Swisse Agric 31:17–22
6. Karoui R, Baerdemaeker JD (2007) A review of the analytical methods coupled with chemometric tools for the determination of the quality and identity of dairy products. Food Chem 102:621–640
7. Andreotti G, Lamanna R, Trivellone E, Motta A (2002) ^{13}C NMR spectra of TAG: an easy way to distinguish milks from different animal species. J Am Oil Chem Soc 79(2):123–127
8. Jha SN, Matsuoka T (2004) Detection of adulterants in milk using near infrared spectroscopy. J Food Sci Technol 41(3):313–316
9. Kasemsumran S, Thanapase W, Kiatsoonthon A (2007) Feasibility of near-infrared spectroscopy to detect and to quantify adulterants in cow milk. Anal Sci 23:907–910
10. Reid LM, O'Donnell CP, Downey G (2006) Recent technological advances for the determination of food authenticity. Trends Food Sci Technol 17(7):344–353
11. Sacco D, Brescia MA, Sgaramella A, Buccolieri A, Ogrinc N, Sacco A (2009) Discrimination between Southern Italy and foreign milk samples using spectroscopic and analytical data. Food Chem 114:1559–1563
12. Nicolaou N, Xu Y, Goodacre R (2010) Fourier transform infrared spectroscopy and multivariate analysis for the detection and quantification of different milk species. J Dairy Sci 93:5651–5660
13. Chen R-K, Chang L-W, Chung Y-Y, Lee M-H, Ling Y-C (2004) Quantification of cow milk adulteration in goat milk using high-performance liquid chromatography with electrospray ionisation mass spectrometry. Rapid Commun Mass Spectrom 18:1167–1171
14. Gutierrez R, Vega S, Diaz G, Sanchez J, Coronado M, Ramirez A, Perez J, Gonzalez M, Schettino B (2009) Detection of non-milk fat in milk by gas chromatography and linear discriminant analysis. J Dairy Sci 92(5):1846–1855
15. Anguita G, Martin R, Garcia T, Moralez P, Haza AI, Gonzalez I, Sanz B, Hernandez PE (1997) A competitive enzyme-linked immunosorbent assay for detection of bovine milk in ovine and caprine milk and cheese using a moloclonal antibody against bovine β-casein. J Food Prot 60(1):64–66
16. Hurley IP, Coleman RC, Ireland HE, Williams JHH (2004) Measurement of bovine IgG by indirect competitive ELISA as a means of detecting milk adulteration. J Dairy Sci 87:543–549
17. El-Rady AA, Sayed M (2006) Identification of milk source by polymersae chain reaction-restriction fragment length polymorphism analysis. J Rapid Methods Autom Microbiol 14:146–155
18. Maccabiani G, pavoni E, Tilola M, Agnelli E, Simoni M, D'Ambrosa F (2005) Setting-up a PCR-based method for species identification in milk products. Vet Res Commun 29:327–329
19. Addeo F, Nicolai MA, Chianese L, Moio L, Musso SS, Bocca A, Giovine LD (1995) A control method to detect bovine milk in ewe and water buffalo cheese using immunoblotting. Milchwissenchaft 50:83–85
20. Lee CC, Chang H-S, Sheen H-S (2004) A quick novel method to detect the adulteration of cow milk in goat milk. Asian Australas J Anim Sci 17(3):420–422
21. Recio I, Garcia-Risco MR, Amigo L, Molina E, Ramos M, Martin-Alvarez PJ (2004) Detection of milk mixtures in Haloumi cheese. J Dairy Sci 87:1595–1600
22. Dias LA, Peres AM, Veloso ACA, Reis FS, Vilas-Boas M, Machado AASC (2009) An electronic tongue taste evaluation: identification of goat milk adulteration with bovine milk. Sens Actuators B 136:209–217

23. Yu H, Wang J, Xu Y (2007) Identification of adulterated milk using electronic nose. Sens Mater 19(5):275–285
24. Angeletti R, Gioacchini AM, Seraglia R, Piro R, Traldi P (1998) The potential of matrix-assisted laser desorption/ionisation mass spectrometry in the quality control of water buffalo mozzarella cheese. J Mass Spectrom 33:525–531
25. Bucknall M, Fung K, Duncan M (2002) Practical quantitative applications of MALDI-TOF spectrometry. J Am Soc Mass Spectrom 13(9):1015–1027
26. Cozzolino R, Passalacqua S, Salemi S, Malvagna P, Spina E, Garozzo D (2001) Identification of adulteration in milk by matrix-assisted laser desorption/ionisation time-of-flights mass spectrometry. J Mass Spectrom 36:1031–1037
27. Fanton C, Delogu G, Maccioni E, Podda H, Seraglia R, Traldi P (1998) Matrix-assisted laser desorption/ionisation mass spectrometry in the dairy industry 2. The protin fingerprint of ewe cheese and its application to detection of adulteration by bovine milk. Rapid Commun Mass Spectrom 12:1569–1573
28. Noo M, Tollenaar R, Ozalp A, Kuppen P, Bladergroen M, Eilers P, Deedler A (2005) Reliability of human serum protein profiles generated with c8 magnetic beads assisted MALDI-TOF mass spectrometry. Anal Chem 77(22):7232–7241
29. Ross P, Hall I, Haff I (2000) Quantitative approach to single peptide polymorphism analysis using MALDI-TOF mass spectrometry. Biotechniques 29(3):620–626, 628–629
30. Soeryapranata E, Powers JR, Hill HHJ, SiemsIII WF, Al-Saad KA, Weller KM (2002) Matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry method for the quantification of β-casein fragment (f193-209). J Food Sci 67(2):534–538
31. Liland KH, Mevik B-H, Rukke E-O, Almoy T, Isaksson T (2009) Quantitative whole spectrum analysis with MALDI-TOF MS, Part II: determining the concentration of milk mixtures. Chemom Intell Lab Syst 99:39–48
32. Liland KH, Mevik B-H, Rukke E-O, Almoy T, Skaugen M, Isaksson T (2009) Quantitative whole spectrum analysis with MALDI-TOF MS, Part I: measurement optimization. Chemom Intell Lab Syst 96:210–218
33. Siuzdak G (1996) Mass spectrometry for biotechnology. Academic, London
34. Brereton RG (2005) Chemometrics: data analysis for the laboratory and chemical plant. Wiley, West Sussex
35. Goodacre R, Timmins EM, Burton R, Kaderbhai N, Woodward AM, Kell DB, Rooney PJ (1998) Rapid identification of urinary tract infection bacteria using hyperspectral whole-organism fingerprinting and artificial neural networks. Microbiology 144:1157–1170
36. Hotelling H (1936) Relations between two sets of variates. Biometrika 28:312–377
37. Johnson RA, Wichern DW (2007) Applied multivariate statistical analysis. Prentice Hall, Upper Saddle River
38. Martens H, Naes T (1989) Methods of calibration. In: multivariate calibration, 1st edn. Wiley, Chichester, pp 73–236
39. Shawe-Taylor J, Christianini N (2004) Kernel methods for pattern analysis. Cambridge University Press, Cambridge
40. Amigo L, Recio I, Ramos M (2000) Genetic polymorphism of ovine milk proetins: its influence on technological properties of milk—a review. Int Dairy J 10:135–149
41. Borkova M, Snaselova J (2005) Possibilities of different animal milk detection in milk and dairy products—a review. Czech J Food Sci 23(2):41–50
42. Moioli B, Pilla F, Tripaldi C (1998) Detection of milk protein genetic polymorphisms in order to improve dairy traits in sheep and goats: a review. Small Rumin Res 27:185–195
43. Recio I, Perez-Rodriguez M-L, Ramos M, Amigos L (1997) Capillary electrophoresis analysis of genetic variants of milk proteins from different species. J Chromatogr A 768:47–56
44. Visser S, Slangen CJ, Lagerwerf FM, Dongen WDV, Haverkamp J (1995) Identification of a new genetic variant of bovine β-casein using reversed-phase high-performance liquid chromatography and mass spectrometric analysis. J Chromatogr A 711:141–150
45. Catinella S, Traldi P, Pinelli C, Dallaturca E, Marsillo R (1996) Matrix-assisted laser desorption/ionisation mass spectrometry in milk science. Rapid Commun Mass Spectrom 10:1629–1637
46. Goodacre R, Neal MJ, Kell DB (1994) Rapid and quantitative analysis of the pyrolysis mass spectra of complex binary and tirtiary mixtures using multivariate calibration and artificial neural networks. Anal Chem 66:1070–1085