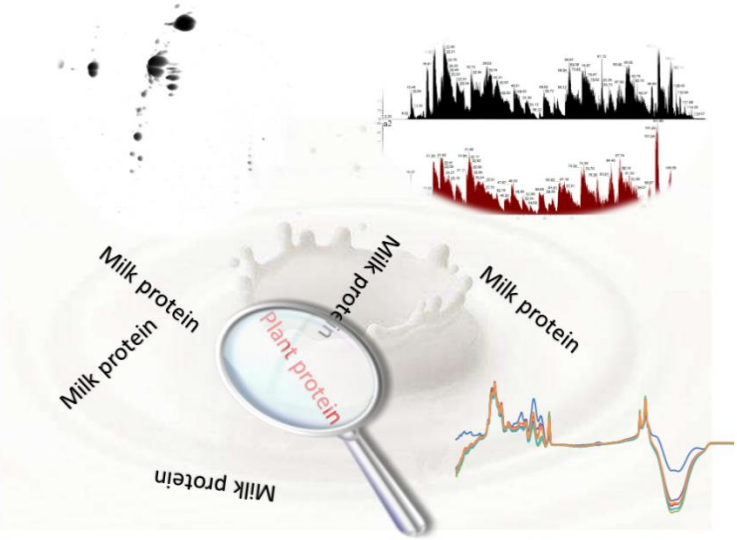
LIÈGE université
Gembloux
Agro-Bio Tech

Milk protein adulteration detection via mass spectra and infrared spectroscopy



Jinhui Yang

COMMUNAUTÉ FRANÇAISE DE BELGIQUE
UNIVERSITÉ DE LIÈGE – GEMBLOUX AGRO-BIO TECH

MILK PROTEIN ADULTERATION DETECTION VIA MASS SPECTRA AND INFRARED SPECTROSCOPY

Jinhui YANG

Dissertation originale présentée en vue de l'obtention docteur en sciences agronomiques et ingénierie biologique

Promoteur ULiège: Hélène SOYEURT
Co-Promoteur: Vincent BAETEN
Promoteur Cotutelle: Jiaqi WANG
Année civile: 2019

This work is supported by:

Résumé

Jinhui Yang (2019). Détection de l'adultération du lait par des sources protéiques via la spectrométrie de masse et infrarouge (Thèse de doctorat) Gembloux, Belgique, Université de Liège, Gembloux Agro-Bio Tech, 157 p., 24 tables, 15 figs.

L'adultération protéique du lait concerne l'ajout dans le lait ou les produits laitiers de substances exogènes riches en azote et présentant un faible coût afin de masquer la teneur naturellement faible en protéine du lait. Les protéines végétales sont de bons candidats pour une telle adultération. Ainsi, les protéines de soja, pois, blé et riz sont les plus prisées par les fraudeurs. La consommation de lait ainsi frelaté peut provoquer des problèmes de santé chez l'humain. Par conséquent, la mise au point de méthodes de détection de ces adultérants dans le lait revêt une grande importance pour garantir la sécurité alimentaire. La présente thèse vise à tester et comparer différentes méthodes comme l'électrophorèse bidimensionnelle sur gel (2-DE), la spectrométrie de masse en tandem avec la chromatographie en phase liquide et la spectroscopie proche et moyen infrarouge pour détecter l'ajout de protéines exogènes dans le lait, particulièrement des protéines végétales hydrolysées. Les principaux résultats sont :

(1) Par les spots de protéines mis en évidence sur le gel de polyacrylamide de lait frelaté, la β-conglycinine et la glycinine ont pu être détectées dans du lait frelaté avec des protéines de soja, tandis que la détection de viciline et de conviciline a indiqué l'addition de protéines de pois. La présence sur le gel de la β-amylase de de la serpine a indiqué quant à elle la présence de protéines de blé dans le lait. Un profil protéique établi par la méthode 2-DE a permis d'identifier le lait adultéré par des protéines de soja et de pois avec une limite de détection de 4% de protéines végétales dans la protéine totale.

(2) Les gels de type dodécylsulfate de sodium et de polyacrylamide (SDS-PAGE) ont clairement révélé qu'une centrifugation à 20 000 g pendant 60 minutes réduisait l'intensité de la bande de caséine et d'albumine dans le lait. Aucun spot protéique n'a été observé pour les protéines hydrolysées de blé et de riz. La nano-HPLC-MS / MS a isolé les principales protéines du soja (β-conglycinine, glycinine), du pois (viciline, conviciline, légumine) et de blé (gluténine et gliadine) dans des laits frelatés. Cette méthode a permis ainsi la détection de laits frelatés par des protéines hydrolysées de soja et de blé avec une limite de détection au-dessus de 0,5% en protéines totales. La limite pour la détection du pois était de 2 et 4%. Aucune protéine de riz n'a pu être identifiée. La nano-HPLC-MS/MS combinée à l'analyse en composantes principales (ACP) a discriminé tous les échantillons adultérés du lait authentique.

(3) La spectroscopie infrarouge à transformée de Fourier utilisant la réflectance totale atténuée (ATR-MIR) appliquée à un lait commercial écrémé a aussi été testée pour identifier l'adultération protéique et pour estimer le contenu en nitrogène non-protéique (eNPN). Des clusters spectraux ont été observés par l'ACP en fonction du contenu et du type d'adulterant. La régression des moindres carrés partiels (PLS) basée sur l'ATR-MIR a montré une bonne performance à prédire eNPN (R^2 = 0.70, RMSE = 0.06 g/100 g obtenus par validation croisée). L'addition d'adultérant protéique dans le lait a bien augmenté le eNPN qui était compris entr 0.04 et 0.37

g/100g. Sur base de la différence entre le range contenu en NPN d'un lait standard et la valeur prédite de eNPN des laits frelatés, il a été possible de détecter des laits adultérés par des protéines hydrolysées de riz (HRP), soja (HSP) et blé (HWP) avec une teneur supérieure ou égale à 3.8 g/L. Cette étude ne peut pas être généralisée car les changements de composition naturels du lait n'ont pas été pris en compte. Par conséquent, une seconde étude a été menée pour valider ces observations et étende l'utilisation de la spectroscopie infrarouge au proche infrarouge.

(4) Ainsi, 9 échantillons de lait de vache ont été frelatés avec HRP (2.5-40 g/L), HWP (1.875-30 g/L), du lactosérum (1.875-30 g/L), de l'urée (0.5-8 g/L) et de l'eau (3.125-50 g/L). Ces 234 échantillons ont été analysés par des spectromètres ATR-MIR et NIR. L'analyse discriminante par PLS sur base du NIR ou ATR-MIR n'a pas permis d'isoler les échantillons de lait de référence. Les spectres ATR-MIR discriminaient mieux l'adultération par HRP et HWP au-dessus de 6.25% alors que les spectres NIR après dérivée première détectaient le lactosérum à partir de 12.5%. Une bonne performance de prédiction a été constatée pour quantifier le contenu en HRP, en urée dans le lait par ATR-MIR et NIR (R^2 de validation > 0.96). ATR-MIR a montré également sa capacité à mieux prédire le contenu en HWP que le NIR (R^2 de validation = 0.95 *vs.* 0.88) alors que le NIR après dérivée première prédisait mieux le lactosérum (R^2 de validation = 0.97 *vs.* 0.40). De plus, les 2 techniques prédisaient bien la teneur totale en protéines exogènes (sans l'eau) dans le lait (R^2 de validation =0.87-0.98 avec un RMSEP=2.04-4.11 g/L). Par conséquent, l'utilisation de méthodes non ciblées est possible pour détecter le niveau de protéines exogènes dans le lait (urée, lactosérum et protéines végétals hydrolysées) en utilisant la spectroscopie NIR ou MIR. Cependant les performances de classification du type d'adultérant ont été plus contrastées selon la méthode et les teneurs en adultérant.

Mots clés: adultération de protéines de lait, électrophorèse bidimensionnelle, spectre de masse tandem par chromatographie en phase liquide à haute performance, spectroscopie proche infrarouge, spectroscopie moyen infrarouge

Abstract

Jinhui Yang (2019). Milk protein adulteration detection via mass spectra and infrared spectroscopy (PhD thesis) Gembloux, Belgium, Univeristy of Liège, Gembloux Agro-Bio Tech, 157 p., 24 tables, 15 figs.

Milk protein adulteration concerns the addition of cheap foreign substance having high nitrogen content into milk and dairy products, to mask low natural protein in milk. Plant protein is a potential candidate of adulterants; soy, pea, wheat, and rice proteins being the most popular. The consumption of those adulterants could induce serious Human health disorders. Therefore, the development of detection methods for protein milk adulteration is of great importance to guarantee food safety. The current thesis aims to test and compare methods such as two-dimensional gel electrophoresis (2-DE), liquid chromatography tandem mass spectrometry, and infrared spectroscopy to detect foreign protein spiked in milk, especially hydrolyzed plant proteins. Here are the major results:

(1) According to the protein spots highlighted on the polyacrylamide gel of adulterated milk, β-conglycinin and glycinin were detected in milk adulterated with soy protein, while legumin, vicilin, and convicilin indicated the addition of pea protein, and β-amylase and serpin marked wheat protein. 2-DE-based protein profile allowed to identify milk spiked with soy and pea protein, with a detection limit of 4% plant protein in the total protein.

(2) Sodium dodecyl sulfate -polyacrylamide gel electrophoresis (SDS-PAGE) gels revealed clearly that centrifugation at 20 000 g for 60 min reduced band intensity of casein and albumin in milk. No obvious protein line was observed for hydrolyzed wheat and rice protein. Results of nano-HPLC-MS/MS highlighted the major proteins of soy (β-conglycinin, glycinin), pea (vincilin, convicilin, legumin) and wheat (glutenin and gliadin) in adulterated milks. So, this method allows the detection of hydrolyzed soy and wheat protein at the level above 0.5% in total protein, and pea protein at the level of 2 and 4%. No rice protein was identified in milk samples adulterated with hydrolyzed rice protein. Combined with principal component analysis (PCA), nano-HPLC-MS/MS discriminated all the adulterated samples from authentic milk.

(3) Attenuated Total Reflectance Fourier Transform Infrared Spectroscopy (ATR-MIR) applied to commercial skimmed milk was also tested to identify protein milk adulteration and to estimate the non-protein nitrogen content (eNPN). Spectral clusters revealed by PCA depended on the level and type of adulterant. The developed partial least square (PLS) regression showed good performance of ATR-FTIR to predict eNPN (R^2 = 0.70, RMSE = 0.06 g/100 g of full cross-validation). The addition of adulterants to milk increased the eNPN level for all samples and ranged from 0.04 to 0.37 g/100 g. Based on the difference between the known NPN range in normal milk and the predicted eNPN of adulterated samples, it was possible to detect samples adulterated with hydrolyzed rice (HRP), soya (HSP) or wheat (HWP) with a content higher or equal to 3.8 g/L. This study cannot be generalized as the natural milk composition change was not taken into account. Therefore, a second study was

conducted to validate these first observations and to enlarge the use of infrared spectroscopy to near-infrared.

(4) So, 9 raw cow milk samples were adulterated with different levels of HRP (2.5-40 g/L), HWP (1.875-30 g/L), whey (1.875-30 g/L), urea (0.5-8 g/L), and water (3.125-50 g/L). Those 234 samples in total were analyzed using ATR-MIR and near-infrared (NIR) spectrometers. The developed NIR and ATR-MIR PLS - discriminant analysis did not discriminate control milk from adulterated samples. Raw ATR-MIR spectra discriminated better on HRP and HWP adulteration above 6.25%, while first derivative NIR spectra detected whey content above 12.5% in milk. Good prediction performance was observed to quantify the level of HRP and urea in raw milk using both ATR-MIR and first derivative NIR spectra (validation R^2p> 0.96). ATR-MIR showed better prediction on HWP level than NIR (validation R^2p=0.95 *vs.* 0.88), while NIR had a better accuracy in whey level quantification (validation R^2p=0.97 *vs.* 0.40). Moreover, both technics predicted well the total protein adulterant level (without water) in adulterated samples (validation R^2=0.87-0.98, RMSE=2.04-4.11 g/L). Consequently, the use of untargeted quantitative analysis is possible to detect the level of protein adulterants (such as hydrolyzed plant protein, urea or whey) in milk using NIR or MIR infrared spectroscopy, although classification of samples between adulterants types gave contrasted performances depending on the adulterant level and the king of spectroscopy used.

Key words: milk protein adulteration, two-dimensional electrophoresis, high performance liquid chromatography tandem mass spectrum, near- and mid- infrared spectroscopy

Acknowledgements

After years of study, I have completed my PhD research work. Now, I want to express my sincere thanks to everyone who accompanied me on my way to becoming a doctor.

It was my pleasure to study at the Gembloux Agro-Bio Tech (GxABT)-University of Liege (ULg), and I am very lucky to be a student of Prof. Hélène Soyeurt. As my supervisor, she taught me how to express myself accurately and clearly in a logical way for manuscript writing, and her huge efforts on my manuscripts contributed to their final acceptance. She also guided me in how to use R software and in the running of chemometric approaches. Vincent Baeten, my co-supervisor, always gave me his professional and constructive suggestions on my experimental design and summary reports of my results. I am really appreciative of his support during my study in both training sessions and my continuous experiment at the Walloon Agricultural Research Centre (CRA-W). Prof. Jiaqi Wang also gave me lots of guidance on the design of my PhD project. It is my honour to be his PhD student, so that I can continue my studies in the Milk Research Team (MRT). As one of the top experts in dairy nutrition and milk quality in China, many of his opinions have inspired me in my academic studies and choice of career. I also thank comments from the other jury members, Marie-Laure Fauconnier, Yves Beckers, Sabine Danthine, and Georges Lognay, especially suggestion from Georges Lognay on my thesis revision. Meanwhile, I want to express my respect to those who work for the joint PhD program supported by ULg and the Graduate School of the Chinese Academy of Agricultural Sciences (GSCAAS).

As one member of the MRT and a former staff member in the inspection centre, my experiences of learning and working will be some of the best memories of my life. Prof. Nan Zheng always encouraged my studies with passion. Dr. Yongxin Yang gave me copious advice on my experiment. Drs. Songli Li and Fang Wen helped me a lot in improving my detection skills. Drs. Shengguo Zhao and Yangdong Zhang expanded my knowledge on microbiology and dairy processing. Sampling and working with Qiaoyan Ye, Xueyin Qu, Jiemei Zhu, and Hongyuan Dai taught me a lot about field tests. Learning with Li Min, Huaigu Yang, and Yanan Gao always filled me with pleasure. I also remember the help provided by Xinyu Hao and Yongfei Zhu. This limited text could not list all the names in my brain, but I am grateful to them all.

During my stay in Belgium, friendly Belgian colleagues showed their kind hearts to foreign students. Dominique Marchal provided me with answers to my questions in daily life. Anne-Catherine Dalcq surprised and entertained me over dinner, and her sampling during my experiment saved me so much time. Yves Brostaux gave me key points on the programming of R script, Lionel Luxen showed me how to print. I am also very glad to have met other members of SIMa. During my experiment at CRA-W, Nicolas Crasset, Quentin Arnould, Ouissam Abbas, Bernard Lecler, and Olivier Minet helped me a lot with instrument operation, sample scanning, and data export. Meanwhile, Joëlle Haine, Alix Raway, Thierry Henckaerts, and other staff at GxABT should be acknowledged for their efforts in accommodating me.

Thanks to my classmates, the room in which you live is always filled with joy. I am very happy to be with all of you. Lin Li's cooking, Xiaodong Guo's songs, Xiaomei Yang's baby, Hejun Lu's jokes, Peipei Zhang's silence, Bowen Hu's buying, Yingying Zhu's dumplings, Peng Li's calling, Wei Rong's banter, as well as the talents of others, impressed me and drove loneliness away from my foreign life. Of course, I will never forget the company of Xuewei Zhou and Lei Zhang in the office.

Finally, I want to thank my father and mother. Thank you for your patience and waiting. Don't worry and take care of yourselves. I am on my way.

Jinhui Yang

2018

Contents table

List of Tables

List of Figures

List of Abbreviations

(e)NPN	(estimated) non-protein nitrogen
(H)SP	(Hydrolysed) soy protein
(P/M)UFA	(Poly-/Mono-) unsaturated fatty acid
1/2D	First/Second derivative
2DE	Two-dimensional electrophoresis
ALB	Albumin
ANN	Artificial neural networks
ATR	Attenuated Total Reflectance
C(Z)E	Capillary (zone) electrophoresis
C/TP	Crude/Ture protein
CHAPS	3-[(3-cholamidopropyl)-dimethylammonio]-1-propanesulfonate
CN	Casein
CV	Coefficients of variation
DA	Discriminant analysis
DTT	Dithiothreitol
EDTA	Ethylene Diamine Tetraacetic Acid
FA	Fatty acids
FIMS	Flow injection mass spectrometry
FT	Fourier transform
GH	Modified Mahalanobis distance
GMP	Glycomacropeptide
HPLC	High Performance Liquid Chromatography
HRP	Hydrolyzed rice protein
HTT	High throughput transmission
HWP	Hydrolyzed wheat protein
IPG	Immobilised pH gradient
MALDI	Matrix-assisted laser desorption ionization
MIR/NIR	Mid-/Near infrared
MLR	Multiple Linear Regression
MS	Mass spectrometer
MS/MS	Tandem mass spectrometer
MSC	Multiplicative scatter correction
NCBI	National Center of Biotechnology Information

PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffer
PCA	Principal component analysis
PLS	Partial Least Square
PP	Pea protein
Q-TOF	Quadrupole time-of flight
R^2	Coefficients of determination
RCT	Rennet coagulation time
RMSE	Root-mean-square error
S/M/LCFA	Short/medium/long chain fatty acids
SD	Standard deviation
SDS	Sodium dodecyl sulfate
SIMCA	Soft independent modelling taxonomy
SNV	Standard normal variate
SVM	Support Vector Machine
UHT	Ultra-high Temperature
VIP	Variable importance in project
α-LA	α-lactalbumin
β-LG	β-lactoglobulin

1

General introduction

1. Introduction

1.1. Dairy production in China

Milk consumption in China raised in last 20 years. According to the report (http://www.chyxx.com/industry/201710/573395.html), growth rate of dairy product consumption ranged from 5% to 30% from 2000 to 2007. Affected by melamine scandal, milk consumption decreased in 2008 and 2009. After 2010, consumers restored faith on local dairy products, and consumption of dairy products increased slowly at rate of -2-8%. Intake of dairy product per capita in China increased from 28 kg in 2012 to 36 kg in 2016; this is less than 260 kg consumed in developed countries such as America and Germany. Different from major intake (about 76%) of dry and powder dairy products in America, local consumers prefer to consume liquid milk (55%), then milk powder (32%) and dry milk products (13%). Even the average value of milk intake per year varied in different regions: people lived in large- and medium-size cities consumed 31 and 23 kg of liquid milk per capita per year respectively, while residents of small cities and countryside drink less than 15 kg. There is a large demand of milk products for people to improve their dietary input.

Increased with milk consumption, production of dairy product in China increased from 1.5 million ton at beginning of 2010 to 2.8 million ton at the end of 2016 (Figure 1-1A). To meet the demand of consumers, foreign dairy products are imported from New Zealand, Australia, and European Union. The imported volume of dairy product per year increased from 0.90 million tons in 2011 to 1.96 million tons in 2016 (Figure 1-1B). Due to the increasing cost of dairy feeding from 2009, raw milk price increased from 2.7 to 3.1 Yuan/kg in 2010, then stabilized at range of 3.2 and 3.6 Yuan/kg in 2010, 2011, 2015, and 2016. For the drought of weather and prevalence of cow disease in 2013 (http://www.chyxx.com/industry/201803/617800.html), milk price soared from 3.50 to 4.12 in second half year of 2013, then decreased slowly to 3.79 in 2014 (Figure 1-1C).

As encouraged by policies and forced by increasing feed, land, and labour cost, more and more cows are breeding in medium- ($100 < \text{cow number} < 1000$) and large- scaled ($\text{cow number} \leqslant 1000$) intensive farms. From 2012 to 2013, number of small dairy farms (cows number below 100) decreased from 30 to 14, while the ratio of medium- and large- scale farms increased from 89.5% to 94.4%. Milk performance of dairy cows in different scale farms are presented in Figure 1-2. Compared with small farms, medium- and large- scale farms tend to produce milk with higher contents of fat and protein, as well as a lower somatic cell count. Milk yield of individual cow is also higher.

Except melamine scandal in 2008, inferior quality milk powder (protein contents below 1%) also caused 13 death and 200 malnutrition of infants and kids in Anhui province in 2004 (Zhang, 2005). After many food safety affairs, series of laws, standards, and guidance are issued to regulate the production of milk products (Jiang et al., 2018). Moreover, supervision and inspection leading by National Food and Drug Administration has been built in these years, targeted on nutrition, contaminant, mycotoxin, microbiology and other potential risk factors, and the results were open to

public. Inspection results in latest 3 years (2014-2016) showed qualified rate of formula powders for infants and kids exceeded 95% (Jiang et al., 2018).

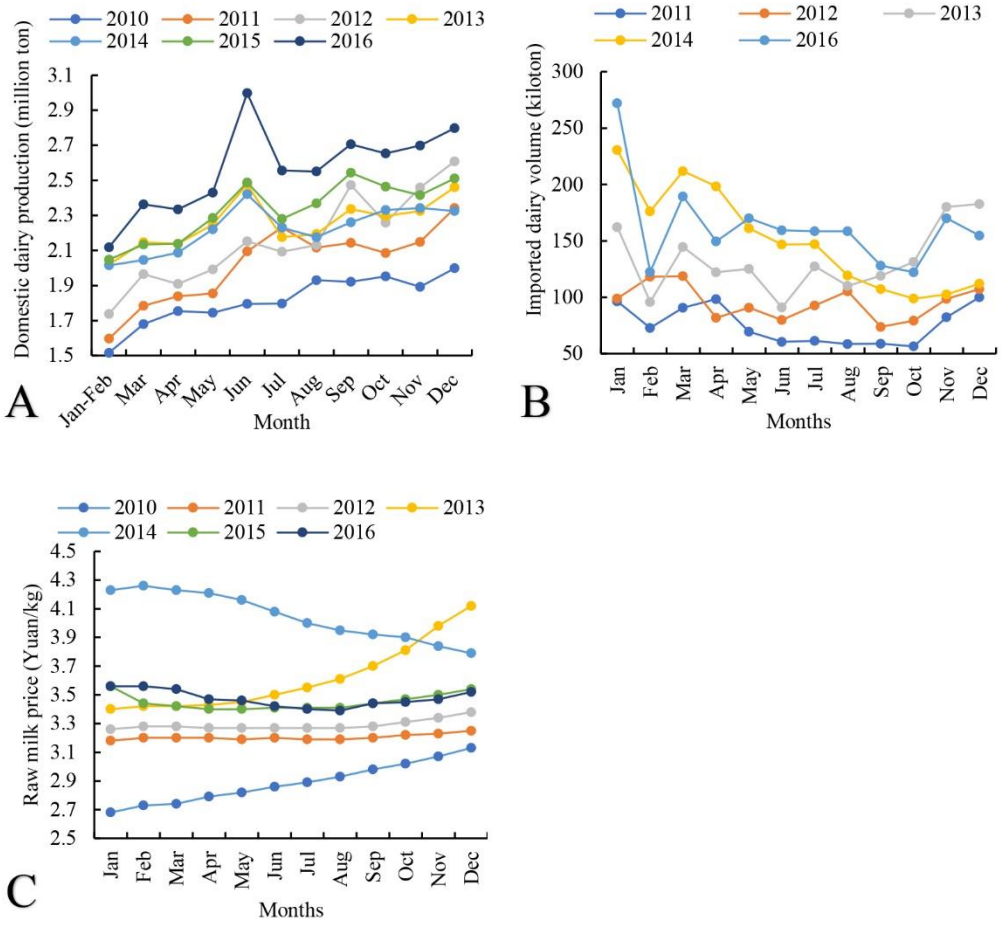


Figure 1-1: Domestic and imported dairy production, and prices of raw milk in each month from 2010 to 2016 in China. Data from China dairy yearbook 2017 (http://www.chinayearbook.com/).

Due to the large gap of cost between foreign and domestic dairy production (http://www.chyxx.com/industry/201803/617800.html), unlabelled reconstitute liquid milk processed from imported powder milk is another top issue in liquid milk authentication (http://www.xinhuanet.com/food/2016-12/06/c_1120059389.htm). In order to prevent the prevalence of reconstitute milk, standard of detection method on derivative of over-heated compounds has been issued by agriculture department (MOA, 2016). Heat-loading of liquid milk has been considered in National Quality Milk Project, which aims to restore consumers' faith on dairy products and promotes dairy production of local farms and industry (Wang, 2012).

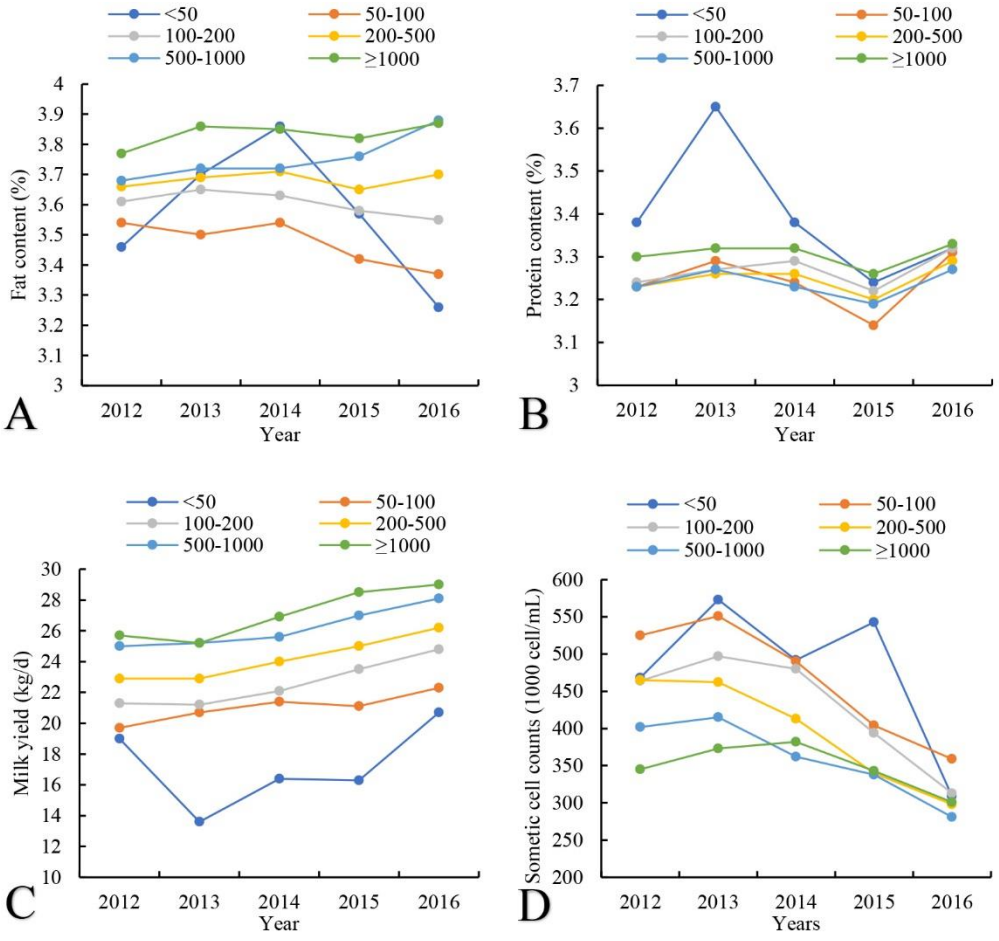


Figure 1-2: Milk performance of dairy cows feed in different sized farms from 2012 to 2016. Data from China dairy yearbook 2017 (http://www.chinayearbook.com/).

1.2. Milk adulteration

To ensure a good and constant quality of milk, the dairy sector fixes the milk price based on some rules as a certain amount of fat, protein, cells and germs. To improve the milk price, the temptation is high to adulterate this "noble" and essential food stuff. However, such adulterated milks are harmful to consumer's health (Nascimento et al., 2017; Poonia et al., 2017). For instance, food safety incidents occurred in 2008 in China, addition of melamine exaggerated protein content in milk and dairy products, which resulted in illness, hospitalization, and death of many children (Domingo et al., 2014).

Generally, milk adulteration involves the addition of cheap, inferior, or even hazardous chemicals or products to increase the volume, mask poor quality, or inflate

nutrients in milk (Nascimento et al., 2017). This practice is common in developing and underdeveloped countries. Indeed, more than 60% of milk in India and around 10% in Brazil is adulterated (Kamal and Karoui, 2015). In order to meet soaring seasonal demand of milk consumption in Pakistan, milk was diluted with water, then added starch, urea, and cane sugar to maintain compositional parameters; meanwhile, formalin, hydrogen peroxide, boric acid and various antibiotics were also added to extend shelf life of products (Afzal et al., 2011). In Brazil, the most frequent contaminants appeared in powdered milk were starch, whey, and sucrose, ranged 20-25%, without obvious flavour changes (Borin et al., 2006). One determination of 300 milk samples in Sudan showed 95% of samples were adulterated with water, and 35.5% with starch (Adam, 2009). An analysis of liquid milk in Kenya found 23.5% of pasteurized and 5.58% of raw samples were positive for H_2O_2, while 23.7% of pasteurized and 19.3% of raw milk were positive for antibiotics (Afzal et al., 2011). A comparative study of milk adulteration in India showed salt, skimmed powder, and urea are the top 3 adulterants, presented in 60-82% of milk samples, while occurrence of other adulterants, such as sucrose, neutralizers, formalin, detergents and H_2O_2, ranged from 22% to 44% (Singuluri and Sukumaran, 2014). Milk adulteration with H_2O_2, gentamycin, and vegetable oil was also reported in China (Salih and Yang, 2017). An adulteration ratio of 30% reconstituted milk in fresh milk was revealed by local media in Taiwan province in China (Salih and Yang, 2017). The most frequent adulterants in milk are summarized in Table 1-1.

Table 1-1: Common milk adulteration practices

Adulterants	objective	Detrimental effect to health	Reference
Water	Increase milk volume	Health concern by contaminated water	(Das et al., 2016; Reddy et al., 2017)
Detergents	Emulsify and dissolve the oil in water	Gastro-intestinal complication, damage to skin and eyes	(Jaiswal et al., 2017)
H_2O_2	Prolong the freshness	Gastritis and inflammation of intestine	(Azad and Ahmed, 2016)
Starch	Increase the non-fat solid	High amounts of addition cause diarrhea	(Singuluri and Sukumaran, 2014)
Na_2CO_3, $NaHCO_3$	Neutralize the acidic effect and preserve milk for long time	Disruption in hormone signaling, abdominal pain, diarrhea, vomiting	(Azad and Ahmed, 2016; Reddy et al., 2017)
Vegetable oil	Increase fat content of milk	-	(Reddy et al., 2017)
Low valued milk	Gain economic profit	Allergy of certain protein	(Jaiswal et al., 2015)

Milk protein adulteration is the result of the addition of nitrogen rich compounds to milk. This leads to an increase of apparent protein content and masks the natural low content of milk protein. Melamine, urea, and whey are the most common protein adulterants (Nascimento et al., 2017). Besides hazard effect of melamine on consumer health, the other two adulterants have also detrimental effects. Excessive urea in milk is harmful to liver and heart, and overburdens kidneys (Kandpal et al., 2012). Addition of rennet whey solid in milk would decrease blood pressure (Reddy et al., 2017). In addition, low price and extensive origin of plant protein products make them attractive as candidate adulterants in milk and dairy products (Haasnoot et al., 2001).

2. Aim and outline of thesis

Plant proteins are used as food additives to improve food characteristics, such as food texture, water retention, and fat emulsification (Garcia et al., 1997). For example, soy protein is often used as non-milk protein in milk replacers, such as simulated yoghurts, coffee whiteners, and frozen desserts (Lopez-Tapia et al., 1999). Besides, soy protein powder is a common adulterant spiked in dairy products, to increase protein content, for its low cost and availability in the market (Poonia et al., 2017). Different preparations of soy protein, such as flours, concentrates, isolates, and hydrolysates, are commercially available (Haasnoot et al., 2001). Compared with other preparations, hydrolysed protein products are inclined to be adulterated in liquid milk, for its higher contents of free amino acids and peptides, as well as higher solubility in water phase (Tessier et al., 2005). Other plant proteins derived from pea, wheat, and rice, are also considered as potential food ingredients (Janssen et al., 1994). Unlabelled addition of these plant protein is not allowed by food labelling regulation (China, 2000; Parliament, 2011). Health concern for consumers on unlabelled addition of plant protein is potential food allergy caused by cupin and prolamin superfamily, such as vicilins and legumins in soybean, and α-amylase and prolamin in wheat respectively (Breiteneder and Radauer, 2004). For example, allergic symptoms of wheat prolamins included atopic dermatitis and exercise-induced anaphylaxis (Shewry et al., 2002). Therefore, it is essential to detect plant protein adulterated in milk.

Series of detection methods have been developed to identify dairy products adulteration with plant protein. In an interlaboratory study, sodium dodecyl sulfate-capillary electrophoresis (SDS-CE) can detect soy and pea protein in milk powder at level of 1-5% in total protein, with standard deviations of repeatability and reproducibility ranged 9-15% and 25-30% respectively (Manso et al., 2002). Calibrated by external soy bean protein isolate as standard, a reversed-phase (RP)-high performance liquid chromatogram (HPLC) method was validated with recovery ranging from 87.9 to 106% in different bovine milks, and the limit of detection was 13 μg/g of samples (Krusa et al., 2000). Chromatographic file of samples collected from HPLC with UV detection (UHPLC) at 215 nm would differentiate samples adulterated with soy, pea, and brown rice protein isolate at 3% and 10% from authentic milk powder (Jablonski et al., 2014). Development of polyclonal antibodies in

immunoassays has made it possible to detect soy, pea, or soluble wheat protein in a range of 1-5% of plant protein in total protein for milk powder adulteration (Haasnoot et al., 2001). Combined with multivariate linear regression (MLR), first derivative NIR absorption showed the best prediction accuracy on soy, pea, and wheat protein isolate in milk powder, with prediction error of 0.23, at range 0-5% of added plant protein (Maraboli et al., 2002). Comparison of these technics is presented in Table 1-2. However, most of these methods focused on milk powder adulteration, while most of adulteration is occurred more likely in liquid milk. Therefore, the aim for this thesis was to compare methods to detect foreign protein adulterated in milk, especially plant protein.

Table 1-2: Summary of detection methods on plant protein in milk adulteration

Techniques	Advantages	Disadvantages	References
Liquid chromatography	Separation of various molecules, identification (with mass spectra) and quantification of target molecules, screen for many compounds	Time consuming, high cost, labor intensive, complex sample pretreatment, need skilled operators	(Kamal and Karoui, 2015)
Immunoassays	Large sample throughput, high sensitivity, low detection limit (0.5%)	Underestimation for UHT treated samples, cross reactivity of antigen, semiquantitative	(Haasnoot and du Pre, 2007; Poonia et al., 2017)
SDS-Capillary electrophoresis	Official detection methods of soy protein, higher accuracy than ELISA	Poor reproducibility, unable to detect soluble wheat protein and soy protein hydrolysates	(Lopez-Tapia et al., 1999; Manso et al., 2002; Sánchez et al., 2002)
Infrared spectroscopy	Fast, low cost, environmentally friendly, rich information, suitable for online quality control, little sample preparation, easy to operate	Low sensibility, low structure selectivity, and overlapped signals and noise bands for NIR spectra; Strong absorption of water for MIR spectra	(Kamal and Karoui, 2015)

Electrophoresis, chromatographic, or spectral fingerprints of food compounds are used to detect milk adulteration (Zhang et al., 2011). Two-dimensional gel electrophoresis (2-DE) would show protein fraction fingerprints of samples on gel map with high resolution, and the separated foreign protein from milk protein could indicate the occurrence of adulteration. High performance liquid chromatography tandem mass spectroscopy (HPLC-MS/MS) exhibits chromatographic fingerprints of peptides in digested samples; peptides from adulterants would be identified by MS.

As a non-invasive and rapid analytical method, infrared spectroscopy presents spectral fingerprints of samples. Combined with chemometrics, absorption difference induced by adulterants is extracted to identify adulteration qualitatively and quantitatively.

This manuscript is a compilation of 3 published and 2 submitted scientific papers. The current chapter (Chapter 1) introduced the dairy context in China as well as the milk adulteration. Chapter 2 will present the potentialities of two-dimensional gel electrophoresis to identify milk adulterated with soy, pea, and wheat proteins. Chapter 3 will show the ability of liquid chromatography–mass spectrometry combined with chemometrics to isolate the differences of peptides between adulterated and control milk. Chapter 4 will review recent advances in milk production and detection by infrared spectroscopy, from composition prediction to quality assessment. Chapter 5 will explore the feasibility of mid-infrared spectroscopy to identify skimmed milk samples spiked with hydrolyzed plant protein (i.e., soy, rice, and wheat) and whey. Chapter 6 will go deeper in the study of the feasibility of infrared spectroscopy to detect milk protein adulteration by using near and mid-infrared rays as well as by increasing the natural variation of protein in milk through the use of individual cow milk samples. Chapter 7 will concern the general discussion of the obtained results. Those will be confronted to the state of art in order to draw a final conclusion and perspectives about the methodology used to detect plan protein adulteration in milk.

3. Thesis of framework

This thesis research is the result of a joint PhD project between Gembloux Agro-Bio Tech-University of Liege (GxABT-ULiège) and Graduate School of Chinese Academy of Agricultural Sciences (CAAS). Funded by Special Fund for Agro-scientific Research in the Public Interest (201403071), Modern Agro-Industry Technology Research System of China (CARS-37, nycytx-04-01), Project of Risk Assessment on Raw Milk (GJFP2016008, GJFP2017008), the Agricultural Science and Technology Innovation Program (ASTIP-IAS12) from Chinese Ministry of Agriculture, researches in this thesis are carried out in Institute of Animal Sciences, CAAS in China before July 25, 2017, and subsequent studies are completed in GxABT-ULiège, with the support of Statistics, Informatics and Applied Modelling Unit (SIMa), Dept. AGROBIOCHEM as well as Food and Feed Quality Unit in Walloon Agricultural Research Centre (CRA-W).

4. Reference

Adam, A. A. H. 2009. Milk adulteration by adding water and starch at Khartoum state. Pakistan Journal of Nutrition 8(4):439-440.

Afzal, A., M. Mahmood, I. Hussain, and M. Akhtar. 2011. Adulteration and microbiological quality of milk (a review). Pakistan Journal of Nutrition 10(12):1195-1202.

Borin, A., M. F. Ferrao, C. Mello, D. A. Maretto, and R. J. Poppi. 2006. Least-squares support vector machines and near infrared spectroscopy for quantification of common

adulterants in powdered milk. Anal Chim Acta 579(1):25-32.
Breiteneder, H. and C. Radauer. 2004. A classification of plant food allergens. The Journal of allergy and clinical immunology 113(5):821-830; quiz 831.
China, N. t. c. o. f. i. o. s. a. o. 2000. General standards for food labelling. Beijing.
Domingo, E., A. A. Tirelli, C. A. Nunes, M. C. Guerreiro, and S. M. Pinto. 2014. Melamine detection in milk using vibrational spectroscopy and chemometrics analysis: A review. Food Res Int 60:131-139.
Garcia, M., M. Torre, M. Marina, F. Laborda, and A. R. Rodriquez. 1997. Composition and characterization of soyabean and related products. Critical Reviews in Food Science & Nutrition 37(4):361-391.
Haasnoot, W. and J. G. du Pre. 2007. Luminex-based triplex immunoassay for the simultaneous detection of soy, pea, and soluble wheat proteins in milk powder. J Agr Food Chem 55(10):3771-3777.
Haasnoot, W., K. Olieman, G. Cazemier, and R. Verheijen. 2001. Direct biosensor immunoassays for the detection of nonmilk proteins in milk powder. J Agr Food Chem 49(11):5201-5206.
Jablonski, J. E., J. C. Moore, and J. M. Harnly. 2014. Nontargeted detection of adulteration of skim milk powder with foreign proteins using UHPLC-UV. J Agric Food Chem 62(22):5198-5206.
Janssen, F., J. De Baaij, and G. Hägele. 1994. Heat-treated meat products. Detection of modified gluten by SDS-electrophoresis, western-blotting and immunochemical staining. Fleischwirtschaft (Germany).
Jiang, H., H. Tian, D. Tian, and X. Ren. 2018. Quality status of infant formula milk powder in recent three years and related regulatory requirements (in Chinese). China Dairy Industry 46(4):39-42.
Kamal, M. and R. Karoui. 2015. Analytical methods coupled with chemometric tools for determining the authenticity and detecting the adulteration of dairy products: A review. Trends Food Sci Tech 46(1):27-48.
Kandpal, S., A. Srivastava, and K. Negi. 2012. Estimation of quality of raw milk (open & branded) by milk adulteration testing kit. Indian Journal of Community Health 24(3):188-192.
Krusa, M., M. Torre, and M. L. Marina. 2000. A reversed-phase high-performance liquid chromatographic method for the determination of soya bean proteins in bovine milks. Anal Chem 72(8):1814-1818.
Lopez-Tapia, J., M. R. Garcia-Risco, M. A. Manso, and R. Lopez-Fandino. 1999. Detection of the presence of soya protein in milk powder by sodium dodecyl sulfate capillary electrophoresis. J Chromatogr A 836(1):153-160.
Manso, M. A., T. M. Cattaneo, S. Barzaghi, C. Olieman, and R. Lopez-Fandino. 2002. Determination of vegetal proteins in milk powder by sodium dodecyl sulfate-capillary gel electrophoresis: interlaboratory study. J Aoac Int 85(5):1090-1095.
Maraboli, A., T. M. P. Cattaneo, and R. Giangiacomo. 2002. Detection of vegetable proteins from soy, pea and wheat isolates in milk powder by near infrared spectroscopy. J near Infrared Spec 10(1):63-69.
MOA, C. 2016. Identification of reconstituted milk in pasteurized and UHT milk. in Standard of MOA. Vol. NY/T 939-2016. The Ministry of Agriculture of the People's

Republic of China, Beijing.
Nascimento, C. F., P. M. Santos, E. R. Pereira, and F. R. P. Rocha. 2017. Recent advances on determination of milk adulterants. Food Chem 221:1232-1244.
Parliament, E. 2011. Regulation (EU) 1169/2011 of the European Parliament and of the Council of 25 October 2011 on the provision of food information to consumers. Off J Eur Communities. L 304:18.
Poonia, A., A. Jha, R. Sharma, H. B. Singh, A. K. Rai, and N. Sharma. 2017. Detection of adulteration in milk: A review. International Journal of Dairy Technology 70(1):23-42.
Reddy, D. M., K. Venkatesh, and C. V. S. Reddy. 2017. Adulteration of milk and its detection: A review. IJCS 5(4):613-617.
Sánchez, L., M. D. Pérez, P. Puyol, M. Calvo, and G. Brett. 2002. Determination of vegetal proteins in milk powder by enzyme-linked immunosorbent assay: Interlaboratory study. J Aoac Int 85(6):1390-1397.
Salih, M. A. M. and S. Yang. 2017. Common Milk Adulteration in Developing Countries Cases Study in China and Sudan: A Review. Advances in Dairy Research 5(4):1-4.
Shewry, P., F. Beaudoin, J. Jenkins, S. Griffiths-Jones, and E. Mills. 2002. Plant protein families and their relationships to food allergy. Portland Press Limited.
Singuluri, H. and M. Sukumaran. 2014. Milk adulteration in Hyderabad, India-a comparative study on the levels of different adulterants present in milk. Journal of Chromatography & Separation Techniques 5(1):1.
Tessier, B., M. Schweizer, F. Fournier, X. Framboisier, I. Chevalot, R. Vanderesse, C. Harscoat, and I. Marc. 2005. Prediction of the amino acid composition of small peptides contained in a plant protein hydrolysate by LC–MS and CE–MS. Food Res Int 38(5):577-584.
Wang, J. 2012. The future strategic direction of dairy industry is to develop quality milk (in Chinese). China animal husbandry and veterinary medicine 39(6):1-5.
Zhang, J. 2005. Four major food safety incidents in 2004 (in Chinese). China Anti-counterfeiting (2):12-13.
Zhang, J., X. D. Zhang, L. Dediu, and C. Victor. 2011. Review of the current application of fingerprinting allowing detection of food adulteration and fraud in China. Food Control 22(8):1126-1135.

2

Detection of plant protein adulterated in fluid milk using two-dimensional gel electrophoresis combined with mass spectrometry

In this chapter, two-dimensional gel electrophoresis was used to separate proteins of adulterated and control milk. Application of immobilized pH gradient (IPG) strips distributed protein spots linearly based on isoelectric points, and protein ladder of reference marker was used to calculate molecular weight in the second direction. Based on the spots difference presented in the gel, the remarkable proteins were identified by MALDI-TOF MS. Compared to control milk, the minimum level of adulterated samples distinguishable on gel maps by scanning was defined as detection limit.

From Yang, J., N. Zheng, Y. Yang, J. Wang, and H. Soyeurt. 2018. Detection of plant protein adulterated in fluid milk using two-dimensional gel electrophoresis combined with mass spectrometry. Journal of food science and technology, 55(7): 2721-2728.

Abstract

The illegal or unlabelled addition of plant protein in milk can cause serious anaphylaxis. For sustainable food security, it is therefore important to develop a methodology to detect non-milk protein in milk products. This research aims to differentiate milk adulterated with plant protein using two-dimensional gel electrophoresis (2-DE) coupled with mass spectrometry. According to the protein spots highlighted on the gel of adulterated milk, β-conglycinin and glycinin were detected in milk adulterated with soy protein, while legumin, vicilin, and convicilin indicated the addition of pea protein, and β-amylase and serpin marked wheat protein. These results suggest that a 2-DE-based protein profile is a useful method to identify milk adulterated with soy and pea protein, with a detection limit of 4% plant protein in the total protein.

Key words: milk adulteration; plant protein; two-dimensional gel electrophoresis; mass spectrometry

1. Introduction

Milk adulteration with exogenous nitrogen-rich components leads to increased apparent protein content, which is used, along with fat content, to define the price of milk. Due to their low cost and abundant sources, vegetable proteins are potential adulterants for dairy products (Haasnoot et al., 2001). Compared with whey, plant proteins in food such as soy protein and wheat gluten would induce a lower postprandial insulin response in consumers (Chalvon-Demersay et al., 2017). However, such plant proteins are also identified food allergens, which can result in hypotension and anaphylaxis (Nakamura and Teshima, 2013). The undeclared addition of these proteins into milk may therefore cause serious health risks. Consequently, the development of analytical methods to detect vegetable proteins in milk products is of paramount importance.

In recent years, a variety of targeted analytical methods have been used to detect the addition of plant proteins (mainly soy, pea, and wheat) in dairy products. Most recent studies of milk adulteration with plant protein have focused on milk powder. Added soy, pea, and soluble wheat protein in milk powder could be detected through immunological tests. Development of polyclonal antibodies in immunoassays has made it possible to detect these adulterants in a range of 1-5% of plant protein in the total milk protein content, although the results seem to be significantly affected by ultra-high temperature (UHT) treatment applied to the samples (Haasnoot et al., 2001; Sanchez et al., 2002). Compared with Enzyme-Linked Immunosorbent Assay (ELISA), the results of vegetable protein percentages obtained from Sodium Dodecyl Sulfate-Capillary Electrophoresis (SDS-CE) had a higher rate of accuracy. However, due to its poor reproducibility, SDS-CE requires improvement in terms of instrument and operation standardisation before large-scale use is feasible (Sanchez et al., 2002). The application of mass spectroscopy (MS) allowed the identification of plant protein added to milk powder. With tetraborate- ethylene diamine tetraacetic acid (EDTA) extraction and trypsin digestion, the peptides from soy and pea proteins present in skimmed milk powder could be identified by quadrupling time-of-flight MS followed by High Performance Liquid Chromatography (HPLC) (Luykx et al., 2007). Untargeted MS was also developed for screening soy and pea protein mixed into skimmed milk powder. The comparative Liquid Chromatography-MS approach enabled unequivocal discrimination between skimmed milk powder (SMP) containing 5% soy or pea protein and unadulterated skimmed milk powder (Cordewener et al., 2009).

Only a few studies have developed detection of the adulterated plant proteins in fluid milk. Attenuated Total Reflectance-Fourier Transform Infrared (ATR-FTIR) has been employed to detect the presence of soy milk in cow's milk. Based on the infrared absorption of peaks of amide, α-34 tocopherol, and soybean kunitz trypsin inhibitor, the spectra of soy milk and control milk adulterated with soy milk presented significant differences to that of control milk (Jaiswal et al., 2015). Calibrated by the external standard method, a reversed phase HPLC method was validated to be good enough in terms of detection performance (such as robustness, reproducibility, accuracy, and precision). The reversed phase HPLC method was used to quantify soy

protein in milk, and had both a rapid separation run (11 min) and low detection limit (13μg/g of bovine milk) (Krusa et al., 2000).

Thanks to the application of high-resolution spectrometers and bioinformatic tools, two-dimensional gel electrophoresis (2-DE) is used to separate protein mixtures in proteomic studies (Pomastowski and Buszewski, 2014). Combined with isoelectric focusing and SDS-polyacrylamide gel electrophoresis (PAGE) in two vertical directions, 2-DE was expected to separate foreign protein from milk protein in adulterated samples with a low detection limit. When compared with control samples, foreign protein spots in the adulterated samples would be visible in the gel map, which is helpful for the detection of exogenous protein in milk products. Coupled with mass spectrometry, 2-DE gel showed the different protein spot distribution of milk from different species, with several unique spots of casein and whey protein serving as markers to differentiate milk adulteration (Yang et al., 2014). These results mean that this method has the potential to detect vegetable proteins in adulterated milk, given different 2-DE gel maps for the major protein spots of animal and plant sources (Zarkadas et al., 2007; Sirtori et al., 2012; Yang et al., 2014).

We assumed that plant protein in fluid milk would be separated from milk protein in the 2-DE gel, and that the foreign protein spots, identified by MS, would indicate the potential adulteration of milk. The objective of this study is to present the gel map of milk adulteration with soy, pea, and wheat protein at low levels (below 8% of total protein) using 2-DE, and to find the marker proteins that can serve as an indicator, identified via matrix-assisted laser desorption ionisation-time of flight mass spectrometry (MALDI-TOF-MS).

2. Materials and methods

2.1. Sample preparation

Soy protein isolate (Nature's Bounty, Inc, Bohemia, NY, USA), pea protein isolate (LifeTime Nutritional Specialties, Inc. Orange, CA, USA), and wheat protein isolate (Honeyville Food Products, Salt Lake City, UT, USA) were used in this study. Raw cow's milk was obtained from a herd located in Beijing (China). Known amounts of plant protein isolate samples were dissolved in phosphate buffer solution (PBS, 0.1 mol/L, pH 7.2) and vortexed for at least 3 min. Then the mixtures were sonicated for 30 min and stirred by magnetic stirrer overnight. Vegetable protein solutions were prepared after centrifugation at 3000 g for 10 min. The protein contents of the raw milk and vegetable protein solution were determined using the Kjeldahl method (KjelROC Analyzer, Furulund, Sweden). Finally, different amounts of plant protein solution were added to raw milk in order to comprise 2%, 4%, and 8% of the total protein. Skimmed milk samples were prepared by centrifugation at 3000 g for 10 min to remove milk fat and stored at -20 ºC until further electrophoresis separation.

2.2. Separation by 2-DE

The protein concentration in thawed samples was determined using bicinchoninic acid (BCA) assay kits (P0010S, Beyotime Institute of Biotechnology, China) before

isoelectric focusing. Samples of a total of 250 mg protein mixed with rehydration buffer (8 M urea, 2 M thiourea, 4% (w/v) 3-[(3-cholamidopropyl)-dimethylammonio]-1-propanesulfonate (CHAPS), 65 mM dithiothreitol (DTT), 0.4% immobilised pH gradient (IPG) buffer and trace bromophenol blue) were loaded onto 17 cm pH 4-7 IPG strips (Bio-Rad), as described by Yang et al. (2014). Isoelectric focusing was carried out at 20°C. The IPG strips were rehydrated overnight and a series of focusing steps were performed as follows: desalting at 50 V for 2 h, 50-100 V for 30 min, 100-500 V for 1 h, 500-1000 V for 1 h, 1000-9000 V for 5 h, and then 9000 V for 80,000 V•h. Before the second separation, IPG strips were immersed in 2% (w/v) dithiothreitol, 0.05 mol/L Tris-HCl pH 8.8, 6 mol/L urea, 30% (v/v) glycerol, and 2% (w/v) sodium dodecyl sulphate (SDS) and shaken gently at room temperature for 12 min, followed another incubation in 2.5% (w/v) iodoacetamide, 0.05 mol/L Tris-HCl pH 8.8, 6 mol/L urea, 30% (v/v) glycerol, and 2% (w/v) SDS for 12 min. Subsequently, strips were transferred to 12% polyacrylamide gels and sealed with 0.5% (w/v) low-melting-point agarose. The electrophoresis conditions used were 50 V for 30 min and 220 V to the end. Then the gels were stained with 0.12% Coomassie Brilliant Blue G-250 solution overnight, and destained with distilled water. Each sample was repeated three times. The gel images were scanned using a GS800 calibrated densitometer (Bio-Rad, USA) and exported to PDQuest 8.0 (Bio-Rad, Hercules, CA, USA) for protein spot analysis. To compare the differences between gels, protein spots were automatically matched and manually compiled, with 'all or none' as the determining criterion. Finally, protein spots detected only in adulterated milk, when compared with pure milk, were selected.

2.3. In-gel digestion, protein identification, and database search

As reported in the existing literature (Yang et al., 2014), the selected protein spots were cut manually from the gels and washed three times in acetonitrile/water (v/v, 50:50). After de-staining and drying, the chopped pieces were incubated with a volume of digestion buffer containing 5 ng sequence-grade trypsin solution for 20 h at 37 °C. Finally, the digestion was stopped through the addition of 100 μL 60% (v/v) acetonitrile in 0.1% (v/v) trifluoroacetic acid. The digested samples were then analysed using a 5800 Plus MALDI TOF Analyzer (Applied Biosystems, Foster City, CA, USA). Protein identification was performed using MASCOT (Matrix Science) to search the uniprot database (http://www.uniprot.org/). PDQuest 8.0 (Bio-Rad Laboratories, CA, USA) was employed to find the unique spots that appeared in adulterated milk in contrast to pure milk.

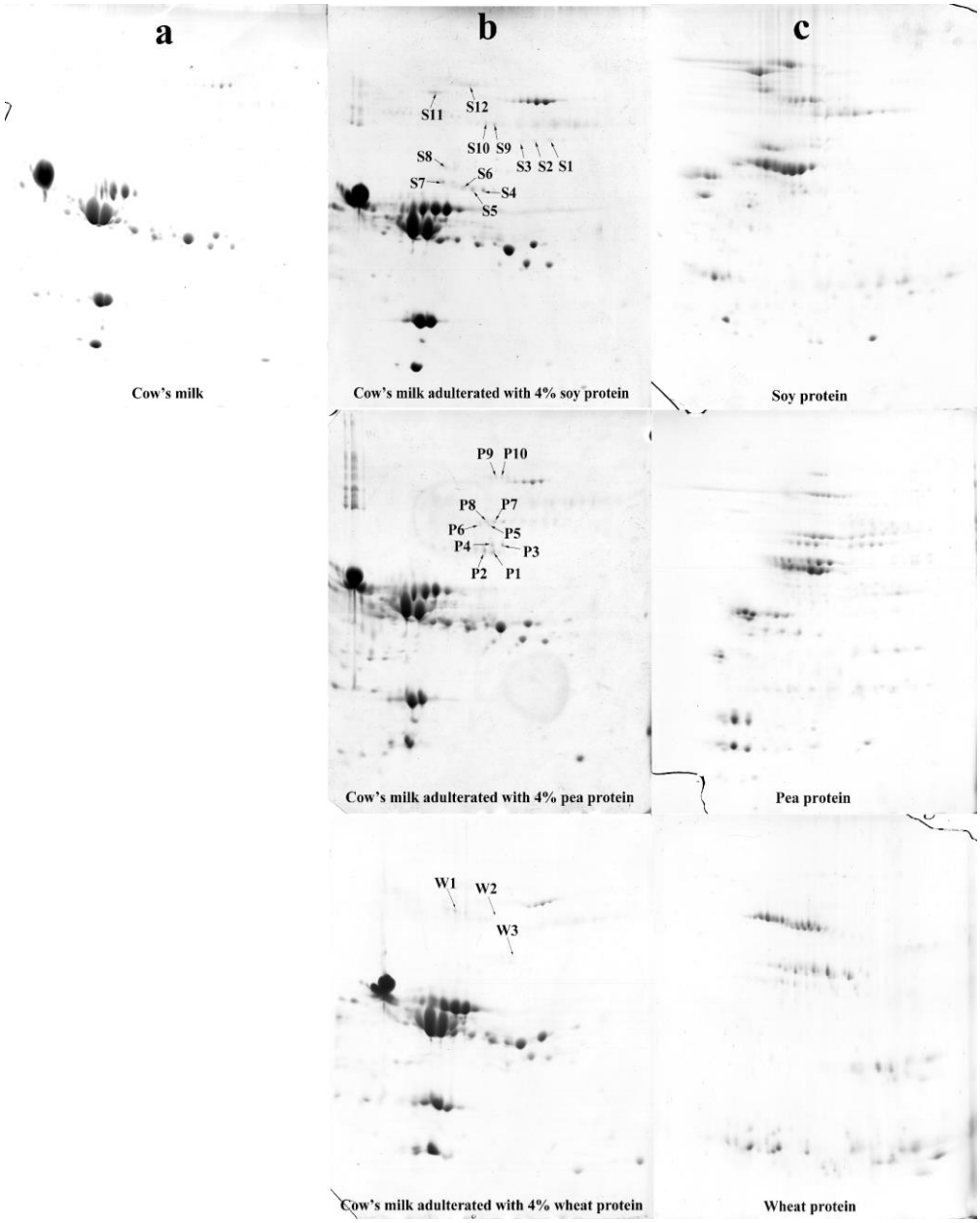


Figure 2-1: Two-dimensional electrophoresis maps of protein spots of cow's milk (a), cow's milk adulterated with 4% plant protein (b), and plant protein (c). Plant protein sources are soy, pea, and wheat. Protein spots labelled with arrow are only detected in adulterated milk compared with cow's milk

Table 2-1: Identification and optical density of protein spots only detected in spiked milk compared with control milk on the gels, analysed by Matrix-Assisted Laser Desorption/Ionisation Time of Flight Mass Spectrometry (MALTI-TOF MS)

Protein spot	Protein name *(organism)*	Uniprot IDs [a]	Molecular mass (kDa) [a]	Isoelectric point [a]	Score	Number of matched peptides	Average relative intensity	CV of relative intensity (%)
S1	β-conglycinin β subunit *(Glycine max)*	F8WQS1	50.010	6.14	644	26	72.1	16.5
S2	β-conglycinin β subunit *(Glycine max)*	F7J077	50.468	5.88	592	26	105.4	37.4
S3	β-conglycinin β subunit (*Glycine max*)	Q50JD8	48.358	5.67	483	24	30.8	17.5
S4	Uncharacterized protein (*Glycine max*)	I1LST1	88.639	5.10	6	40	246.6	15.3
S5	Uncharacterized protein (*Glycine max*)	I1L939	71.254	5.23	3	43	254.2	6.8
S6	Uncharacterized protein (*Medicago truncatula)*	G7I2I6	6.749	5.14	3	47	211.1	12.1
S7	Glycinin G2 (*Glycine soja*)	A0A0B2PSP9	59.640	5.79	350	14	287.0	8.4
S8	Glycinin A3B4 subunit (*Glycine max*)	Q7GC77	58.608	5.52	132	10	96.4	23.5

Table 2-1 Continued

Protein spot	Protein name (organism)	Uniprot IDs [a]	Molecular mass (kDa) [a]	Isoelectric point [a]	Score	Number of matched peptides	Average relative intensity	CV of relative intensity (%)
S9	β-conglycinin α subunit (*Glycine max)*	Q94LX2	70.549	5.12	628	20	94.4	55.2
S10	β-conglycinin α subunit (*Glycine max)*	O22120	63.184	4.92	725	23	87.9	55.7
S11	β-conglycinin α subunit (Glycine soja)	A0A0B2Q6W9	70.521	5.09	262	18	345.0	2.8
S12	β-conglycinin α prime subunit (*Glycine max)*	Q4LER6	72.469	5.50	233	22	244.8	5.3
P1	Legumin A (*Pisum sativum*)	P15838	59.633	6.21	295	15	361.3	8.2
P2	Legumin A (*Pisum sativum*)	Q9T0P5	59.153	6.16	137	12	387.7	6.2
P3	Legumin A (*Pisum sativum*)	P15838	59.633	6.21	305	11	214.4	10.9
P4	Legumin A (*Pisum sativum*)	P15838	59.633	6.21	241	14	253.4	5.5

Table 2-1 Continued

Protein spot	Protein name (organism)	Uniprot IDs [a]	Molecular mass (kDa) [a]	Isoelectric point [a]	Score	Number of matched peptides	Average relative intensity	CV of relative intensity (%)
P5	P54 protein (*Pisum sativum*)	O49927	55.027	6.05	445	18	60.4	51.3
P6	P54 protein (*Pisum sativum*)	O49927	55.027	6.05	521	18	57.5	18.5
P7	Vicilin (*Pisum sativum*)	P13918	52.257	5.39	211	18	128.8	9.6
P8	Vicilin (*Pisum sativum*)	P13918	52.257	5.39	670	26	116.2	24.2
P9	Convicilin (*Pisum sativum*)	Q9M3X6	72.134	5.50	427	23	65.1	3.2
P10	Convicilin (*Pisum sativum*)	Q9M3X6	72.134	5.50	643	26	81.4	13.3
W1	β-amylase (*Triticum aestivum)*	W5EKI0	61.360	5.00	938	16	228.4	20.4
W2	β-amylase (Triticum aestivum)	W5C8P9	57.105	5.29	368	6	97.1	9.8

Table 2-1 Continued

Protein spot	Protein name (organism)	Uniprot IDs a	Molecular mass (kDa) a	Isoelectric point a	Score	Number of matched peptides	Average relative intensity	CV of relative intensity (%)
W3	Serpin 3 (Triticum aestivum)	C0LF32	43.227	5.56	1010	13	97.8	38.3

a ID, Molecular mass and isoelectric point derived from the uniport database.

3. Results and discussion

In the current study, 2-DE maps were constructed for cow's milk, soy, pea, and wheat plant protein extracted by PBS, and cow's milk adulterated with 2%, 4%, and 8% soy, pea, and wheat protein. The well-resolved and repeatable 2-DE gel maps of cow's milk, plant proteins, and milk adulterated with plant protein at the 4% level are presented in Figure 2-1. There were no significant differences between raw milk and milk adulterated with plant protein at the 2% level. Typical vegetable protein spots were observed for milk adulterated with 4% and 8% pea or soy protein, with only a few foreign protein spots being detected in milk adulterated with 4% and 8% wheat protein. The foreign protein spots detected only in adulterated milk were selected as marker proteins. The relative intensity and identification of these spots is listed in Table 2-1.

Due their potential to be food allergens (Nakamura and Teshima, 2013), the detection of plant proteins in adulterated milk is an important matter for public health. Therefore, the hypothesis of this research was to use 2-DE as a method to differentiate plant proteins from milk proteins, with a low detection limit (2% plant protein), in fluid skimmed milk spiked with soy, pea or wheat protein. For that, two sub-objectives must be fulfilled. Firstly, the 2-DE method must detect the presence of plant proteins in adulterated milk, and secondly, this detection must be reliable.

3.1. Sample preparation

The solubility of vegetable protein in skimmed milk and its availability for isoelectric focusing electrophoresis produced a good separation of plant protein via the 2-DE method in this study. PBS was used to prepare a plant protein solution in our experiment. Moderate pH values in PBS promoted the dissolution to aqueous solution of β-conglycinin and glycinin from soy-source protein, and legumin and vicilin from pea-source protein (Samoto et al., 2007). Tetraborate-EDTA buffer was an effective tool for extracting plant protein from milk powder, as 2% of milk protein, 94% of soy-source protein, and 87% of pea-source protein were retrieved in the pellet after centrifugation (Luykx et al., 2007; Scholl et al., 2014). Moreover, protein extraction at pH 8.3 via tetraborate-EDTA buffer did not support isoelectric electrophoresis in this study (Scholl et al., 2014).

3.2. Detection of plant protein

Protein spots which were only detected in milk adulterated with soy protein were labelled S1-S12 (Fig. 3-1, Table 3-1). Gel maps containing soy protein showed soy protein spots related mainly to β-conglycinin and glycinin. β-conglycinin contained α (such as spots S9 to S12) and β (such as spots S1 to S3) subunits. Their identified molecular weight ranged from 63 to 72 kDa and from 48 to 50 kDa, respectively. Their isoelectric points ranged from 4.92 to 5.50 and from 5.67 to 6.14 for α and β subunits, respectively. The molecular weight and isoelectric point of glycinin identified in this study (spots S7 to S8) were between 59 and 60 kDa and between 5.52 and 5.79. Our results were similar to those previously reported (Zarkadas et al.,

2007). Although three additional spots detected in soy protein adulterated milk (spots S4 to S6) were not characterised by the Uniprot database, they have a location comparable to acidic subunits of glycinin observed by Zarkadas et al. (2007). Based on their acidic isoelectric point, spots S7 and S8 were likely to also be acidic subunits of glycinin. Disulfide bonds linking the subunits of glycinin were broken by the DDT used in the sample preparation and, therefore, the acidic subunits with a molecular weight of about 35 kDa would be dissociated from the basic subunits (Nishinari et al., 2014). This explains why the molecular weight of spots S4 to S8 identified in our 2-DE gels was around 36-37 kDa. Results for the optical intensity for the spots S4 to S7 and S11 to S12, the relative density with a value above 200, and coefficients of variation (CV) ranging from 2.8% to 15.3%, suggest high β-conglycinin and glycinin content in soy protein adulterated milks. This was confirmed in a study by Zarkadas et al. (2007), where α and α' subunits of conglycinin, as well as acidic subunits of glycinin, accounted for 21-33% and 35-45% of the soy protein content. The differentiation of soy protein spots compared to milk protein spots can be explained by the difference in molecular weight between these different proteins. Indeed, the identified soy protein spots (for example, S1 to S4 and S7 to S12) have a higher molecular weight than casein and whey protein, whose molecular weight is below 35 kDa. In conclusion, β-conglycinin and glycinin are identified in this study as the indicator of milk adulteration with soy protein. Peptides originating from these proteins were also detected by mass spectrometry in reports on milk powder adulteration (Luykx et al., 2007; Cordewener et al., 2009).

Protein spots (P1-P10) detected in milk spiked with pea protein (Fig. 3-1), were identified as legumin A, vicilin, convicilin, and P54 protein from peas in the uniprot database (Table 3-1). Pea protein spots identified as having higher relative intensities (from 214 to 387) were P1-P4. Their CV ranged from 55% to 10.9%. The relative intensity and CV for the rest of the identified pea protein spots (P5-P10) were 60% to 129% and 3.2% to 51.3%, respectively. The molecular weight of these proteins varied from 52 to 72 kDa and their isoelectric point ranged from 5.39 to 6.21. The higher molecular weight of these proteins compared to milk proteins explains the observed separation on 2-DE gels. Meanwhile, it was observed that some of the extracted pea protein fractions were covered by casein and whey protein and therefore were not visible in the 2-DE gels of adulterated milks. The observed high relative intensity of spots P1 to P4 and the moderate relative intensity of spots P7 and P8 suggest high legumin contents and moderate vicilin contents in pea protein. This accords with previous findings (Sirtori et al., 2012), which also stated that the soluble globulin in pea seeds accounted for approximately 70% of the total protein; the two major proteins were vicilin and legumin, with a minor protein, convicilin. Tzitzikas et al. (2006) suggested that the ratio of legumin to vicilin ranged from 2 to 4 in pea globulin. Similar to glycinin in soy protein, legumin subunits (around 60 kDa) consisted of one acidic α (35-43 kDa) and one basic β (19-23 kDa) polypeptide, linked together via a disulfide bridge (Gatehouse et al., 1980). Legumin, identified in this study, was also found to be a major protein in kidney beans (Parmar et al., 2014). Based on their location on the gel, spots P1 to P4 in our 2-DE gels were more likely to be the acidic α polypeptides of legumin. P54 protein detected in our study was mentioned by Wang

et al. (2012) as an important storage protein in peas. As the major seed protein of peas, peptides from legumin and vicilin were also detected in adulterated milk powder (Luykx et al., 2007; Cordewener et al., 2009).

The gel map of milk adulteration with wheat protein highlighted W1-W3 as marker proteins (Figure 3-1), which were characterised as beta-amylase and serpin 3 from wheat in the uniprot database. Fewer spots with high intensities for significant proteins were observed in wheat protein adulterated milk compared with those observed in soy and pea protein adulterated milks. This may pose some difficulty for unequivocal discrimination. The identified β-amylase (W1, W2) and serpin (W3) were shown to be wheat-source protein fractions in research performed by Becker et al. (2012). Only spot W1 indicated the existence of wheat protein in adulterated milk with high intensity (228.4), the other two spots showed low optical density (97). These spots also have a higher molecular weight (43-61 kDa) than major milk proteins. The wheat protein fraction extracted by PBS in this study was inconsistent with KCl soluble/methanol-soluble fractions from wheat flour in previous reports, due to the absence of gliadins or glutenin subunits (Hurkman and Tanaka, 2004). The protein in wheat seed is comprised of glutenins, gliadins, albumins, and globulins (Hurkman and Tanaka, 2004). The insolubility of gliadin and aggregation of glutenin were likely to contribute to the absence of these proteins in the supernatant after centrifugation (Becker et al., 2012). Interestingly, analysis of pellets of skimmed milk powder adulterated with wheat protein isolate after tetraborate-EDTA extraction has also failed to identify wheat source proteins (Scholl et al., 2014).

3.3. Reliability

The separation of soy and pea protein adulteration in milk was more visible than in the case of wheat protein adulteration. Typical foreign protein spot groups for pea and soy proteins appeared in the gel map at a 4% level of adulteration. Marker proteins (S4-S7, S11, S12, P1-P4) with high intensity (211.1-387.7) could be considered as indicators of milk adulteration and showed good reliability (CV ranged from 2.8% to 15.3%). In contrast, only one spot (W1) detected for 4% wheat protein adulterated milk had an intensity above 200. Therefore, 2-DE used in our study was of limited use to detect the presence of wheat protein in milk.

In this study, the detection limit of 2-DE for soy and pea protein in milk adulteration was 4% (8 μg). No obvious soy and pea protein spot appeared in the 2% plant protein adulterated samples, and only appeared when the foreign protein level was more than 8 μg, as found in the 4% plant protein adulterated milk. Yang et al. (2014) observed around 1μg of bovine α-lactalbumin and β-lactoglobulin in a gel of goat's and camel's milk samples adulterated with 2% cow's milk; however, this limit detection was not confirmed in this study. The poor detection limit found in the present study can be ascribed to the incomplete aqueous solubility of plant protein. In future research, low non-target protein preparation (ultracentrifugation) may improve the resolution and detection limit for plant protein with 2-DE (Yang et al., 2014). In addition, the application of fluorescence staining could allow a lower limit (1-2 ng) of protein detection than Coomassie Brilliant Blue staining (8-16 ng) for 2-DE detection, as

applied in current study (Berggren et al., 2000).

2-DE separated soy protein β-conglycinin and glycinin, pea protein legumin, vicilin, and convicilin, as well as wheat protein β-amylase and serpin from milk casein and whey protein on the gel of adulterated milk. This not only highlights the exogenous plant protein from milk protein, but prevents the masking of the trace amounts of plant protein by highly abundant casein and whey proteins in the successive identification by MS. Preliminary screening for manufactured skimmed milk powder containing 5% soy or pea protein isolates using MS showed few plant protein peptides were identified (Luykx et al., 2007). Similar peak profiles were also observed in the comparative LC-MS analysis between skimmed milk powder and skimmed milk powder adulterated with 5% soy protein isolates, and discrimination of these samples needs multivariable analysis by a post-alignment clustering procedure (Cordewener et al., 2009). From this point of view, 2-DE provided high sensitivity and specificity for the final identification of added plant protein, using MALDI-TOF MS.

4. Conclusions

The results demonstrate that 2-DE would be effective for screening milk adulterated with at least 4% soy and pea protein, with successive mass spectrometry analysis identifying several peptides: β-conglycinin and glycinin from soy and legumin, vicilin and convicilin from peas. For milk adulteration with wheat protein, only β-amylase and serpin were identified. As a potential detection method, 2-DE is robust for the validation of milk adulteration with soy and pea protein. Moreover, an improved electrophoresis procedure with special sample preparation and staining methods would reduce the limit of detection effectively.

5. Acknowledgement

This study was financially supported by Agro-scientific Research in the Public Interest (201403071), the Agricultural Science and Technology Innovation Program (ASTIP-IAS12), Modern Agro-Industry Technology Research System of the PR China (CARS-37), and the Project of Risk Assessment on Raw Milk (GJFP2016008), granted by Ministry of Agriculture (PRC).

6. References

Becker, D., H. Wieser, P. Koehler, A. Folck, K. H. Muhling, and C. Zorb. 2012. Protein composition and techno-functional properties of transgenic wheat with reduced alpha-gliadin content obtained by RNA interference. J Appl Bot Food Qual 85(1):23-33.

Berggren, K., E. Chernokalskaya, T. H. Steinberg, C. Kemper, M. F. Lopez, Z. Diwu, R. P. Haugland, and W. F. Patton. 2000. Background-free, high sensitivity staining of proteins in one- and two-dimensional sodium dodecyl sulfate-polyacrylamide gels using a luminescent ruthenium complex. Electrophoresis 21(12):2509-2521.

Chalvon-Demersay, T., D. Azzout-Marniche, J. Arfsten, L. Egli, C. Gaudichon, L. G.

Karagounis, and D. Tome. 2017. A Systematic Review of the Effects of Plant Compared with Animal Protein Sources on Features of Metabolic Syndrome. J Nutr 147(3):281-292.
Cordewener, J. H. G., D. M. A. M. Luykx, R. Frankhuizen, M. G. E. G. Bremer, H. Hooijerink, and A. H. P. America. 2009. Untargeted LC-Q-TOF mass spectrometry method for the detection of adulterations in skimmed-milk powder. J Sep Sci 32(8):1216-1223.
Gatehouse, J. A., R. D. Croy, and D. Boulter. 1980. Isoelectric-focusing properties and carbohydrate content of pea (Pisum sativum) legumin. Biochemical Journal 185(2):497-503.
Haasnoot, W., K. Olieman, G. Cazemier, and R. Verheijen. 2001. Direct biosensor immunoassays for the detection of nonmilk proteins in milk powder. J Agr Food Chem 49(11):5201-5206.
Hurkman, W. J. and C. K. Tanaka. 2004. Improved methods for separation of wheat endosperm proteins and analysis by two-dimensional gel electrophoresis. J Cereal Sci 40(3):295-299.
Jaiswal, P., S. N. Jha, A. Borah, A. Gautam, M. K. Grewal, and G. Jindal. 2015. Detection and quantification of soymilk in cow-buffalo milk using Attenuated Total Reflectance Fourier Transform Infrared spectroscopy (ATR-FTIR). Food Chem 168:41-47.
Krusa, M., M. Torre, and M. L. Marina. 2000. A reversed-phase high-performance liquid chromatographic method for the determination of soya bean proteins in bovine milks. Anal Chem 72(8):1814-1818.
Luykx, D. M. A. M., J. H. G. Cordewener, P. Ferranti, R. Frankhuizen, M. G. E. G. Bremer, H. Hooijerink, and A. H. P. America. 2007. Identification of plant proteins in adulterated skimmed milk powder by high-performance liquid chromatography-mass spectrometry. J Chromatogr A 1164(1-2):189-197.
Nakamura, R. and R. Teshima. 2013. Proteomics-based allergen analysis in plants. J Proteomics 93:40-49.
Nishinari, K., Y. Fang, S. Guo, and G. O. Phillips. 2014. Soy proteins: A review on composition, aggregation and emulsification. Food Hydrocolloid 39:301-318.
Parmar, N., A. S. Virdi, N. Singh, A. Kaur, R. Bajaj, J. C. Rana, L. Agrawal, and C. S. Nautiyal. 2014. Evaluation of physicochemical, textural, mineral and protein characteristics of kidney bean grown at Himalayan region. Food Res Int 66:45-57.
Pomastowski, P. and B. Buszewski. 2014. Two-dimensional gel electrophoresis in the light of new developments. Trac-Trend Anal Chem 53:167-177.
Samoto, M., M. Maebuchi, C. Miyazaki, H. Kugitani, M. Kohno, M. Hirotsuka, and M. Kito. 2007. Abundant proteins associated with lecithin in soy protein isolate. Food Chem 102(1):317-322.
Sanchez, L., M. D. Perez, P. Puyol, M. Calvo, and G. Brett. 2002. Determination of vegetal proteins in milk powder by enzyme-linked immunosorbent assay: Interlaboratory study. J Aoac Int 85(6):1390-1397.
Scholl, P. F., S. M. Farris, and M. M. Mossoba. 2014. Rapid Turbidimetric Detection of Milk Powder Adulteration with Plant Proteins. J Agr Food Chem 62(7):1498-1505.
Sirtori, E., I. Isak, D. Resta, G. Boschin, and A. Arnoldi. 2012. Mechanical and

thermal processing effects on protein integrity and peptide fingerprint of pea protein isolate. Food Chem 134(1):113-121.
Tzitzikas, E. N., J. P. Vincken, J. de Groot, H. Gruppen, and R. G. Visser. 2006. Genetic variation in pea seed globulin composition. J Agric Food Chem 54(2):425-433.
Wang, W. Q., I. M. Moller, and S. Q. Song. 2012. Proteomic analysis of embryonic axis of Pisum sativum seeds during germination and identification of proteins associated with loss of desiccation tolerance. J Proteomics 77:68-86.
Yang, Y. X., N. Zheng, J. H. Yang, D. P. Bu, J. Q. Wang, L. Ma, and P. Sun. 2014. Animal species milk identification by comparison of two-dimensional gel map profile and mass spectrometry approach. Int Dairy J 35(1):15-20.
Zarkadas, C. G., C. Gagnon, V. Poysa, S. Khanizadeh, E. R. Cober, V. Chang, and S. Gleddie. 2007. Protein quality and identification of the storage protein subunits of tofu and null soybean genotypes, using amino acid analysis, one- and two-dimensional gel electrophoresis, and tandem mass spectrometry. Food Res Int 40(1):111-128.

3

Detection of plant protein in adulterated milk using non-targeted nano-high performance liquid chromatography - tandem mass spectroscopy combined with principal component analysis

In this chapter, plant protein (soy and pea) and hydrolyzed plant proteins (wheat and rice) were used to prepare the adulterated samples. High-performance liquid chromatography tandem mass spectroscopy was used to separate peptides of control and adulterated samples, with a three-dimensional approach using selected informative data: retention time, mass to charge ratio of precursor ion and product ions. Combined with principal components analysis, the fingerprints of peptides could discriminate all adulterated samples from control milk. Therefore, the detection limit of HPLC-MS/MS in this study was the minimum level of each type of adulterated samples.

From Yang, J., N. Zheng, H. Soyeurt, Y. Yang, and J. Wang. 2018. Detection of plant protein in adulterated milk using non-targeted nano-high performance liquid chromatography -tandem mass spectroscopy combined with principal component analysis. Food Science and Nutrition (available online).

Abstract

The objective of this study was to detect plant protein adulterated in fluid milk using nano-high performance liquid chromatography (HPLC) -tandem mass spectroscopy (LC-MS/MS) combined with proteomics. Unadulterated milk and samples adulterated with soy protein, pea protein, hydrolyzed wheat protein, and hydrolyzed rice protein were prepared, with plant protein level ranged from 0.5% to 8% in total protein. Sodium dodecyl sulfate -polyacrylamide gel electrophoresis (SDS-PAGE) gels clearly revealed that centrifugation at 20 000 g for 60 min would reduce band intensity of casein and albumin in milk. Results of nano-HPLC-MS/MS indicated the major proteins of soy (β-conglycinin, glycinin), pea (vincilin, convicilin, legumin) and wheat (glutenin and gliadin) in adulterated milks, allowing detection of soy protein and hydrolyzed wheat protein at the level above 0.5% in total protein, and pea protein at the level of 2 and 4%. No rice protein was identified in milk samples adulterated with hydrolyzed rice protein. Combined with principal component analysis, nano-HPLC-MS/MS could discriminate all the adulterated samples from authentic milk. This study demonstrated the feasibility of nano-HPLC-MS/MS on the detection of (hydrolyzed) plant protein adulterated in milk.

Key words: milk adulteration, plant protein, sodium dodecyl sulfate - polyacrylamide gel electrophoresis, nano-high performance liquid chromatography - tandem mass chromatography, high speed- centrifugation

1. Introduction

Milk products are considered to be the second highest food in the adulteration database, behind olive oil (Moore et al., 2012). The addition of foreign nitrogenous compounds to milk products to mask original low protein content is common in dairy adulteration (Nascimento et al., 2017). Adulterants in milk products can cause serious food safety incidents, e.g., melamine (Moore et al., 2012). Vegetable protein is a potential candidate to spike milk products for economic reasons (Luykx et al., 2007). Some allergens from plant proteins can cause serious anaphylaxis and disorders (Nakamura and Teshima, 2013), so unlabeled or illegal addition could threaten consumer health and food safety. For these reasons, it is necessary to develop effective techniques to detect plant proteins in milk.

Detection of plant protein in dairy products have been reported in previous literatures. Capillary zone electrophoresis (CZE) has been approved as the official reference method to detect soy protein in skimmed milk powder (Manso et al., 2002). An automated fluorescent microsphere-based flow cytometric triplex immunoassay was developed to detect soy protein (SP), pea protein (PP), and soluble wheat protein in milk powder simultaneously, and the limit of quantification of this triplex immunoassay was above 0.1% (Haasnoot and du Pre, 2007). Detection of soy, pea, wheat, rice protein at 0.1% -0.2% of sample weight in milk powder was realized by a rapid turbidimetric measure based on the absorbance of the resuspended pellet solution (Scholl et al., 2014). Whereas, these methods fail to present the origin of these adulterants. With amino acid sequences revealed by fragmented peptides, mass spectrometry (MS) is successful in the identification of plant protein added to milk products (Luykx et al., 2007; Cordewener et al., 2009; Lu et al., 2017). High performance liquid chromatography (HPLC) -mass spectrometry (MS) can identify numerous peptides from major seed proteins of soy and pea in the adulterated milk powder, after borate buffer extraction and tryptic digestion (Luykx et al., 2007). Although previous studies have shown that borate buffer was effective to extract insoluble soy and pea protein from milk powder (Luykx et al., 2007; Scholl et al., 2014), the borate buffer enrichment step may not be effective in the detection of plant protein in adulterated fluid milk, because soluble foreign protein is dominant in the adulterated protein and should be the target of detection. Hydrolyzed plant protein tends to have high solubility due to its high content of free amino acids and peptides (Aaslyng et al., 1998). A previous study had found that sodium dodecyl sulfate (SDS) -capillary electrophoresis (CE) failed to detect hydrolyzed SP in adulterated milk powder (Lopez-Tapia et al., 1999). Combined with multivariable statistics, a variety of non-targeted detection methods have been proposed to identify plant protein adulterated in raw milk. Partial least squares -discriminant analysis (PLS-DA) and principal component analysis (PCA) using the fingerprints of intact protein flow injection mass spectra (MS) and ultra-high-performance liquid chromatography (UHPLC) -quadrupole time-of-flight (QTOF) -MS were able to detect SP and PP in adulterated milks at the 1% level (Lu et al., 2017; Du et al., 2018). Based on the chromatographic files of authentic and adulterated milk powder obtained by UHPLC with UV detection at 215 nm, the t test approach and multivariate Q statistic from a

SIMCA model would classify milk powder with SP at 1% and 3% levels correctly, and failed to recognize adulterated samples with brown rice and hydrolyzed wheat protein below 10% (Jablonski et al., 2014).

The objective of this study is to identify the (hydrolyzed) plant protein in adulterated milk using non-targeted liquid chromatography-tandem mass spectrometry (LC-MS/MS). PCA is used to reveal the differences of proteins between samples identified by MS. High-speed centrifugation of samples prior to MS is expected to reduce the cover signal from a high abundance of milk protein over small amounts of adulterant protein, and the corresponding separation would be validated by SDS- polyacrylamide gel electrophoresis (PAGE).

2. Materials and methods

2.1. Sample preparation

Pasteurized milk samples were purchased from Sanyuan Foods (Beijing, China). The following plant protein products were used in this study: SP isolate (Nature's Bounty, Inc., Bohemia, NY11716, USA), PP isolate (LifeTime Nutritional Specialties, Inc., Orange, CA92865, USA), HWP (CP100, Conpro, Kangke Food Engineering Tech Ltd., Wuxi, Jiangsu, China) and hydrolyzed rice protein (HRP) (Shuaixing, Yongguodanbaifen Ltd., Wuhan, Hubei, China). About 10 g of plant protein powder was added to 100 ml phosphate buffer (pH = 6.8, 0.2 M). After magnetic stirring overnight, the plant protein solutions were obtained through centrifugation at 5000 g for 20 min followed by filtration with a 0.2 μm syringe filter (13 mm, GHP Minispike, Waters). The protein contents in SP, PP, HWP, HRP solution, and milk were 31.0, 23.5, 52.1, 66.4, and 30.7 mg/mL, respectively, as determined by KjelROC analyzer KD310 (OPSIS AB Inc, Sweden) using the Kjeldahl method (IDF, 2014).

A series of "adulterated" milks (containing 0.5, 1, 2, 4, 8 g of plant protein/100 g total protein) were prepared by mixing the plant solution and milk in mass proportions. Skimmed samples were collected after centrifugation at 5000 g for 20 min. Additional high-speed centrifugation at 20 000 g for 1 h was used to prepare samples before further LC-MS/MS analysis. Both samples (before and after centrifugation at 20 000 g) were analyzed by SDS-PAGE.

2.2. SDS-PAGE

2.2.1. Gel electrophoresis

Sample protein concentrations were spectrophotometrically determined using bicinchoninic acid (BCA) assay kits (P0010S, Beyotime Institute of Biotechnology, China) before analysis. SDS-PAGE were undertaken according to Laemmli (1970). After heating at 95°C for 5 min with an equal volume of 2× SDS-PAGE loading buffer, samples containing 30 mg protein were loaded onto a 12% SDS-PAGE gel, and the separation was performed at 120 V for 2 h. The gels were stained for 8 h in Coomassie blue dye solution [0.12% (w/v) Coomassie brilliant blue G250, 0.12% (w/v) ammonium sulfate, 10% (v/v) phosphoric acid, 20% (v/v) methanol]. This was

followed by destaining steps, in which gels were washed by shaking in 10% (v/v) ethanol and 10% (v/v) acetic acid (destaining solution). Triplicate destained gels were scanned and optically analyzed with Quantity One software (V4.6.2, Bio-Rad, CA, USA). Unique protein bands in the gel of adulterated samples were excised and trypsin digested following the method of Yang et al. (2014).

2.2.2. Protein identification and database search

MS and MS/MS of extracted peptides were collected using a 4800 plus matrix-assisted laser desorption/ionization (MALDI)- time of flight/ time of flight (TOF/TOF) Analyzer (Applied Biosystems, Foster City, CA, USA) equipped with a 355 nm Nd:YAG laser at an acceleration voltage of 20 kV. Acquisition of positive ions was completed in reflector mode by delayed extraction. Peptide masses ranged from 800 to 4000 Da. The top eight precursor ions with a signal-to-noise ratio more than 50 for each sample were processed in tandem MS mode with 2500 laser shots, and collision energy set as 20 keV. The National Center for Biotechnology Information (NCBI) non-redundant database was used to identify the protein via MASCOT (Matrix Science) search. Peaks with a signal-to-noise ratio below 15 were excluded from the search. The search parameters were set as follows: fixed and variable modifications were carbamidomethylation of cysteine and methionine oxidation, tolerance for one missing cleavage, monoisotopic mass accuracy below 100 ppm, fragment and peptide mass tolerances were ± 0.4 Da and ± 100 ppm.

2.3. LC-MS/MS analysis

2.3.1. Protein digestion

100 μL of samples were mixed with same volume of lysis buffer (8 M Urea, 100 mM TrisHCl, pH 8.0), treated by ultrasound (100 W, 10 s, interval 15 s, 10 times) and bathed in ice. After centrifugation at 12 000 g at 4°C for 15 min, supernatants were collected for protein concentration test using a Bradford test (Bio-Rad, Shanghai, China). Then samples containing 200 μg protein were reduced with dithiothreitol (DTT) at a final concentration of 10 mM and incubated at 37°C for 2 h. After cooling to room temperature, samples were mixed with 55 mM iodoacetamide, and vortexed at 600 rpm for 1 min, then incubated at 37°C in the dark for 30 min. The same volume of 100 mM NH_4HCO_3 were added to samples to decrease urea concentration to less than 2 M. Next, 4 μg trypsin was mixed with the samples and kept at 37°C overnight. The digestion was stopped by addition of 100 μL 60% (v/v) acetonitrile in 0.1% (v/v) formic acid solution. StageTip with Empore C18 extraction disks (3M, South Eagan, MN) was prepared to desalt and dry the samples. Authentic milk (control) and samples adulterated with SP and HWP at 0.5-4% were prepared in triplicate, and adulteration with PP and HRP at 2% and 4% levels were prepared in duplicate in this part.

2.3.2. LC-MS/MS analysis

The tryptic digestion products were separated by nano -HPLC prior to Q Exactive HF Mass Spectrometry (Thermo Scientific). The separation conditions were adapted from Cordewener et al. (2009). Samples were injected on a Thermo Scientific EASY column (C18, 2 cm × 100 μm, 5 μm), which was equilibrated with 95% of solvent A

before sample loading, and the peptides were separated on a Thermo Scientific EASY C18 column (100 mm × 75 μm, 3 μm) at a flow rate of 250 nL/min. Solvent A consisted of aqueous 0.1% formic acid solution and solvent B consisted of 84% acetonitrile in aqueous 0.1% formic acid solution. Gradient conditions started at 5% B, then a linear gradient to 8% B at 2 min, then a linear gradient of 23% B at 90 min, then a linear gradient to 40% B at 105 min, then a linear gradient to 100% B at 110 min, and 100% B was maintained for the final 10 min.

Peptide analysis was performed in positive ion mode for 120 min, with a selected mass range of 300 -1 800 mass/charge (*m/z*). For the survey scan, resolving power was set to 60 000 at *m/z* 200, maximum ion injection time was 50 ms, and the automatic gain control target was 3e6. MS/MS data were acquired using the top 20 most abundant precursor ions, as determined by the survey scan, and activation type was HCD. These were selected with an isolation window of 1.5 *m/z* and fragmented via higher energy collisional dissociation with normalized collision energies of 27 eV. For the MS/MS scans, dynamic exclusion of the selected precursor ions was set to 30 s, resolving power was set to 15 000 at *m/z* 200, maximum ion injection time was fixed at 50 ms.

2.4. Data analysis

Raw files were processed by the Maxquant software (version 1.5.3.17) of the selected species database. The protein databases of bovine, soybean, pea, wheat, and rice were downloaded from Uniprot, which contained 138 035, 250 621, 88 489, 393 298, and 753 301 proteins, respectively. The following parameters were applied: trypsin was the enzyme, and two missed cleavages were allowed up, carbamidomethylation of cysteine was defined as a fixed modification; and oxidation of methionine and acetylation of protein N-term were set as variable modifications. Main search and first search of MS/MS ions were set at 6 and 20 ppm, and MS/MS tolerance was 20 ppm. The false discovery rate for protein and peptide identification was 1 %. Relative quantification of identified protein was calculated from the intensities of razor and unique peptides. The decoy database pattern was set as the reverse of the target database.

Identified protein intensities were output to process using Unscrambler software (version 10.4, CAMO AS, Trondheim, Norway). Data processing was described as Cordewener et al. (2009), after log transformation of protein intensities, data standardization before PCA was performed by centering (subtracting median intensities) and normalization (dividing by the standard deviation).

3. Results and Discussion

3.1. SDS-PAGE

Results of SDS-PAGE of skimmed milk samples and samples treated with high speed centrifugation are listed in Figure 3-1 and Figure 3-2, respectively. The distinct bands labelled in Figure 3-1a, Figure 3-1b, Figure 3-2a, and Figure 3-2b were

identified by MALDI-TOF/TOF MS, and the protein information is listed in Table 4-1. As shown in Figure 4-1, the major proteins in skimmed milk consisted of albumin, α-, β-, and κ-casein, β-lactoglobulin, and α-lactalbumin. Several protein bands observed in the lane of SP (Figure 3-1a) and PP (Figure 3-1b) had a similar location to SDS-PAGE data for pea and soy samples reported in a previous study (Scholl et al., 2014), although plant protein extraction methods differed. Although some faint bands are observed between 11 and 17 kD in the lane of HWP (Figure 3-1c), most protein fraction residues from HWP and HRP (Figure 3-1d) are gathered in the bottom line, and this is in line with previous findings, in which 95% of the peptides of hydrolysates were below 1000 Da (Tessier et al., 2005). Similar protein profiles are presented for milk and adulterated milk with 0.5-4% levels of SP, and only the 8% level sample shows weak stripes of β-conglycinin (α and α' subunit, labelled S2 and S1), and glycinin (G2, labelled S3) besides milk protein (Figure 3-1a). Obvious stripes of PP (vicilin and legumin A2, labelled P1 and P2) are observed in milk adulteration at 4-8% level (Figure 3-1b). No visible lane variance appears between milk and samples adulterated with HWP and HRP (Figure 3-1c and Figure 3-1d).

After high speed centrifugation, weak albumin and casein bands appeared for milk protein (Figure 3-2), while increased intensity was observed in plant protein lanes. Decreased milk protein intensities indicate more visible foreign protein lines from plant protein in adulterated milk. Additional protein lines are identified as α subunits of β conglycinin (S4) and glycinin (S5) emerging at 4% and 8% levels of adulteration with SP. Visible S1, S2, and S3 appeared in adulterated samples at levels of 1% (Figure 3-2a). Another protein band (convicilin, labelled P3) from PP could be observed in lanes for 4% and 8% levels of adulteration with PP, and P1 and P2 could be observed at all levels of adulteration with PP (Figure 3-2b). We found still no obvious difference between different levels of adulteration with HWP after high speed centrifugation treatment (Figure 3-2c). Interestingly, as the adulteration level of HRP increased, the intensities of casein and albumin lines was found to decrease (Figure 3-2d), possibly as a result of high NaCl content (40% in dry matter) in hydrolyzed plant protein (Aaslyng et al., 1998). Saturation of milk with NaCl precipitates the casein and albumin while the major whey proteins remain soluble (Fox et al., 2015).

Centrifugation at 5 000 g for 20 min was used to prepare skimmed milk in this study, and ultra-centrifugation at 100 000 g for 1 h could sediment most (90-95%) of the casein from whey (Fox et al., 2015). Therefore, enhanced centrifugation above 5000 g would reduce the amount of casein in milk, and it is confirmed by the comparison of gel electrophoresis (Figure 3-1 vs. Figure 3-2) in current study. The detection limit of SDS-PAGE for SP and PP in milk reduced from 8% (soy) and 4% (pea) to 1% and 0.5%, respectively, and more visible SP and PP (S1-S5, and P1-P3) lines in adulterated samples appeared after centrifugation at 20 000 g at 4°C for 60 min. In other words, high-speed centrifugation for skimmed milk is an alternative pretreatment, which may magnify the minor difference between low amounts of plant protein in adulterated milk revealed by successive LC-MS/MS analysis.

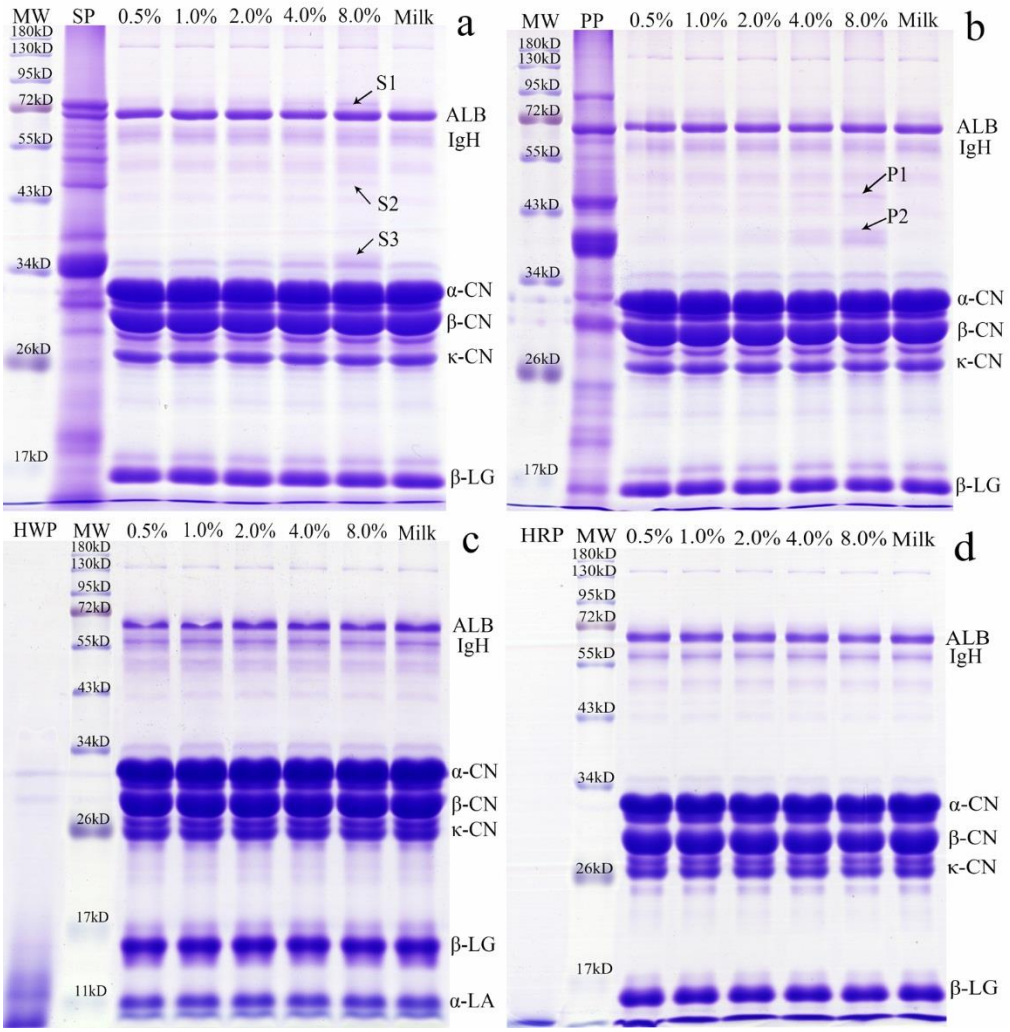


Figure 3-1: SDS-PAGE gel profile of milk adulterated with soy protein (a), pea protein (b), hydrolyzed wheat protein (c), and hydrolyzed rice protein (d), centrifugation at 5000 g for 20 min. SP, Soy protein; PP, pea protein; HWP, hydrolyzed wheat protein; HRP, hydrolyzed rice protein; MW, molecular weight; ALB, albumin; IgH, immunoglobulin heavy chain; CN, casein; α-LA, α-lactalbumin; β-LG, β-lactoglobulin

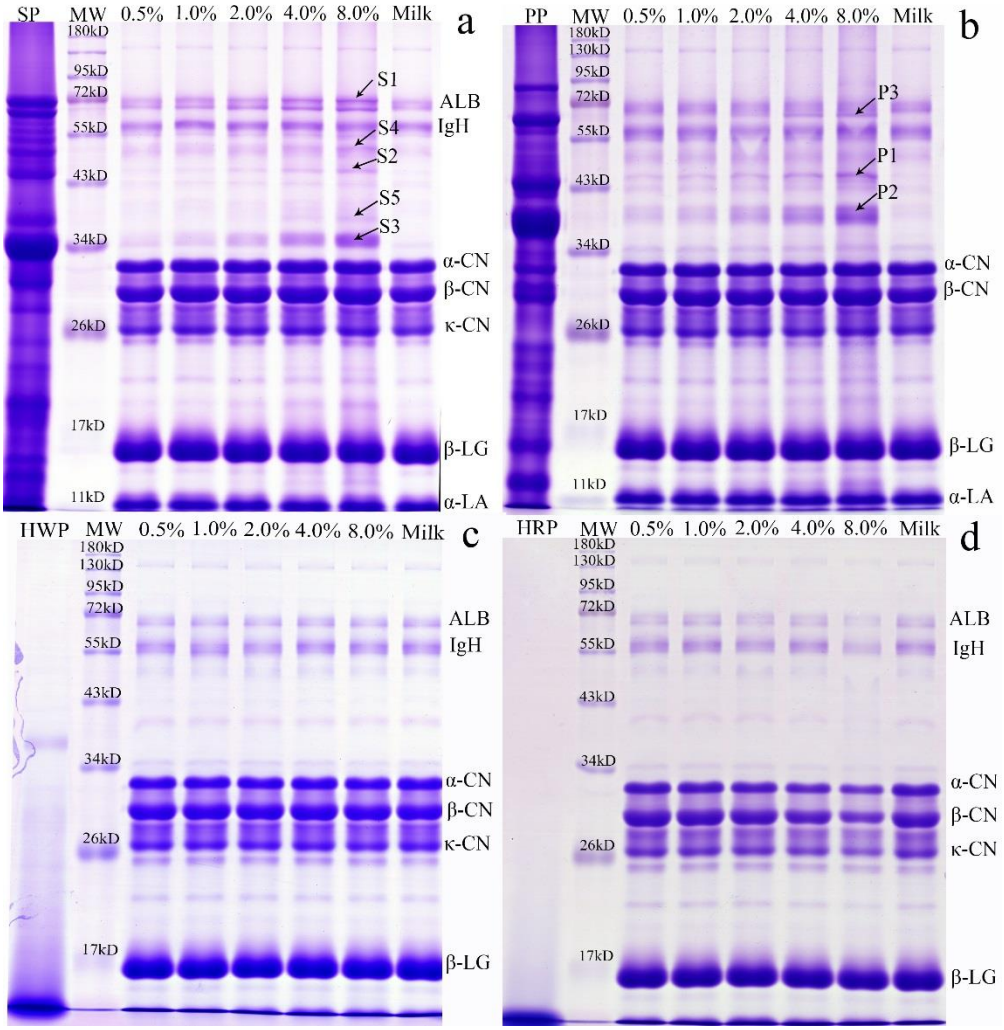


Figure 3-2: SDS-PAGE gel profile of milk adulterated with soy protein (a), pea protein (b), hydrolyzed wheat protein (c), and hydrolyzed rice protein (d), centrifugation at 20 000 g for 60 min. SP, Soy protein; PP, pea protein; HWP, hydrolyzed wheat protein; HRP, hydrolyzed rice protein; MW, molecular weight; ALB, albumin; IgH, immunoglobulin heavy chain; CN, casein; α-LA, α- lactalbumin; β-LG, β-lactoglobulin

Table 3-1: Identification of marker protein spots in adulterated milk contrasted with control milk on the gel by MALDI-TOF MS

Band	ID	Protein name	Organism	Molecular weight (kDa)	Protein isoelectric point	Peptide count	Protein Score
P1	P13918	Vicilin	*Pisum sativum*	52199.7	5.39	25	464
P2	P15838	Legumin A2	*Pisum sativum*	59233.6	6.21	20	315
P3	Q9M3X6	Convicilin	*Pisum sativum*	72019.7	5.50	29	374
S1	Q9FZP9	α' subunit of β-conglycinin	*Glycine max*	65103.4	5.23	31	605
S2	Q94LX2	β-conglycinin α subunit	*Glycine max*	63248.8	5.00	18	520
S3	A0A0B2PSP9	Glycinin G2	*Glycine soja*	59013.1	5.79	14	291
S4	O22120	α subunit of β conglycinin	*Glycine max*	63126.9	4.92	20	576
S5	Q9SB12	Glycinin	*Glycine max*	55337.2	5.46	14	279

3.2. LC-MS/MS coupled with multivariable statistics

The total ion chromatogram of SP, PP, HWP and HRP are shown in Figure 3-S1. There are 430, 902, 356, and 9 proteins identified in SP, PP, HWP, and HRP solutions, respectively. Compared with other plant proteins, fewer peaks appeared in the peptide chromatograms generated from HRP, and fewer proteins were identified. The destruction of tryptophan (Trp) and cysteine (Cys), deamination of glutamine (Gln) and asparagine (Asn), as well as high levels of hydrolysis occurred in the manufacturing process (Aaslyng et al., 1998) may have disturbed the proteomic identification of rice protein in adulterated milk in this study. More adulterant proteins were identified in milk spiked with HWP than samples with HRP. More peptide peaks observed in the chromatogram profile (Figure 3-3d) of HWP indicated less extensive hydrolysis in the manufacturing process for wheat protein, a result which was also confirmed by gel electrophoresis (Figure 3-1c).

Figure 3-S2 shows the summed spectra of replicated measurements for control milk and milk samples adulterated with SP (4% level). No obvious visible difference was observed in the peak intensities of typical LC-MS runs between replicates or among samples, which indicates the reproducibility of sample measurements and the similarity of major peptides between samples. Visible differences between pure milk and samples adulterated with SP at 10% level on MS fingerprints and chromatographic files were observed in recent studies using flow injection MS and ultra HPLC-UV detection respectively (Jablonski et al., 2014; Du et al., 2018). However, direct comparison of chromatograph profiles does not often reveal the difference between adulterated samples and control milk. Discrimination of milk

powder adulterated with 5% SP from control samples by visual inspection of peak profiles was not realized in previous reports, using either HPLC-MS or LC-QTOF MS (Luykx et al., 2007; Cordewener et al., 2009). Therefore, necessary multivariable statistics, such as PCA, should be used to discover the minor difference in peak profiles between adulterated samples and control milk in this study.

Table 3-2: Summary of samples and identified protein number

Item	Adulterant	Level	Replicates	Number of identified proteins	CV of intensities log values (%)1	Percentage of adulterant protein
Milk	None	0	3	418	0.804(0.02-13.8)	0
Soy 0.5	Soy protein isolate	0.5	3	372	1.077(0.00-28.5)	23.4
Soy 1	Soy protein isolate	1	3	403	1.121(0.03-10.8)	29.8
Soy 2	Soy protein isolate	2	3	423	1.052(0.01-19.5)	33.3
Soy 4	Soy protein isolate	4	3	421	0.809(0.01-18.1)	37.8
Pea 1	Pea protein isolate	2	2	272	0.831(0.00-7.00)	20.2
Pea 2	Pea protein isolate	4	2	280	0.672(0.00-14.6)	21.4
Wheat 0.5	Hydrolyzed wheat protein	0.5	3	329	1.392(0.01-14.8)	19.8
Wheat 1	Hydrolyzed wheat protein	1	3	333	1.095(0.06-14.0)	20.7
Wheat 2	Hydrolyzed wheat protein	2	3	337	1.078(0.06-12.1)	22.8
Wheat 4	Hydrolyzed wheat protein	4	3	339	0.790(0.00-17.9)	27.4
Rice 2	Hydrolyzed rice protein	2	2	145	0.813(0.00-6.86)	0
Rice 4	Hydrolyzed rice protein	4	2	145	0.684(0.00-3.68)	0

[1] CV, coefficients of variation, expressed as median (range)

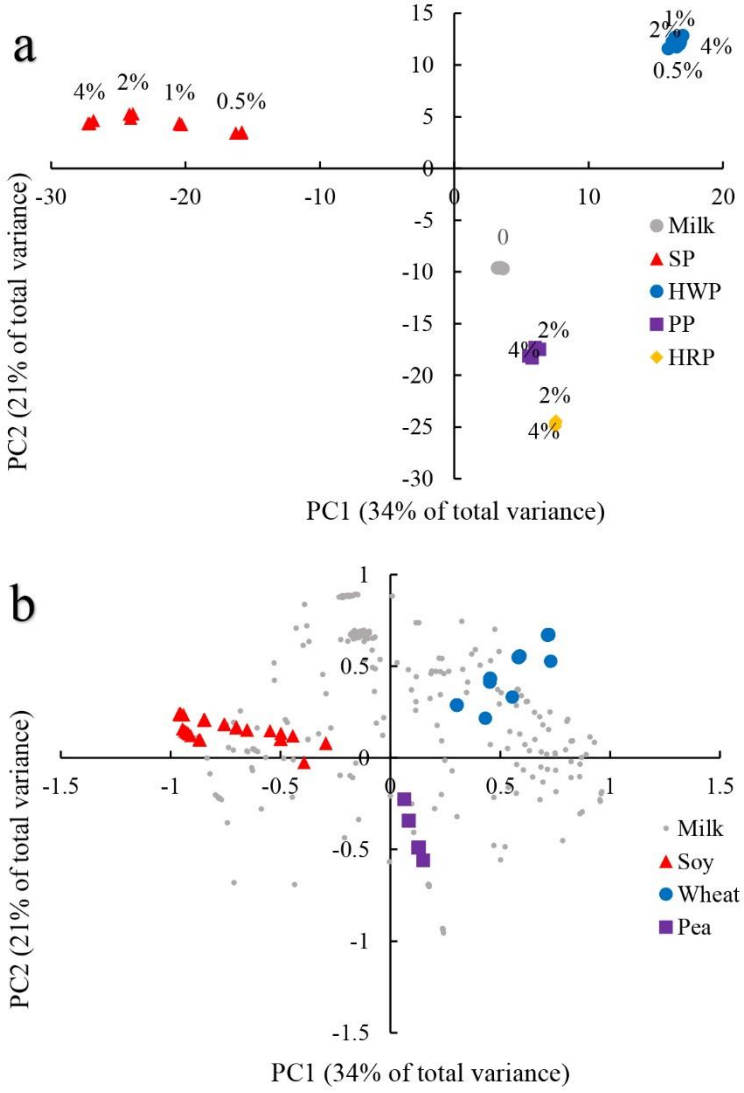


Figure 3-3: Score (a) and correlation loading (b) plots of principal component analysis (PCA) for adulterated and control milk. a, numbers labelled above sample points are the percentage of plant protein in total sample milk protein, and different colors indicate the different adulterated (SP, soy protein; PP, pea protein; HWP, hydrolyzed wheat protein; HRP, hydrolyzed rice protein) or control milk samples. Coloured points in b show the identified protein from adulterants or milk

Score and correlation loading plots of PCA analysis are shown in Figure 3-3. The first 2 PCs accounted for 56% of the total variance, which clearly distinguishes adulterated milk from control milk. In the loading plot (Figure 3-3b), identified SP and wheat protein are separated by PC1 for their opposite loading values; PP are also divided for their negative PC2 loadings, while milk protein scatters evenly across the plot. A score plot (Figure 3-3a) lists each individual LC-MS/MS profile as one point and replicated sample points overlap. Samples adulterated with different plant proteins cluster into four groups and are separated from authentic milk. Samples adulterated with SP, PP, and HWP tend to have a similar location to corresponding identified adulterant proteins in the loading plot. The distance between each level of adulteration with SP and PP is larger than that of adulteration with HWP and HRP. An approximate linear relationship of data points dependent on protein levels could be observed for samples adulterated with SP. Results of PCA in the current study are similar to those reported in other literature (Cordewener et al., 2009; Lu et al., 2017). Our results show that the adulterated samples could be separated from authentic milk for adulterants proteins. Although no rice protein was identified in the samples adulterated with HRP (Table 3-2), these samples were also distinguishable from pure milk.

Descriptive statistics for proteins in samples identified by LC-MS/MS are listed in Table 3-2. Reproducible peak intensities for sample measurements were also presented by coefficients of variation (CV) values, which is ranged from 0.00 to 28.5%, and corresponding medians were below 2%, this was comparable to previous reports (Cordewener et al., 2009). There were 372-421 and 272-280 proteins identified in samples adulterated with SP and PP respectively, while 329-339 and 145 proteins were identified in adulteration with HWP and HRP respectively. About 19.8-37.8% of the total identified protein was found to be adulterant protein from soy, pea, and wheat, and no rice protein was identified. As the adulterant levels increased, the ratio of identified adulterant protein in total protein also increased, except for adulteration with HRP. The top 10 adulterant proteins from adulteration with SP, PP, and HWP are shown in Table 4-S1. Among them, β-conglycinin, glycinin, trypsin inhibitor from SP, and vincilin, convicilin, legumin, provicilin from PP were also identified in other studies, using ultra HPLC-QTOF MS proteomics, as reported by Lu et al. (2017). Meanwhile, proteins identified from SDS-PAGE are also presented in the results of LC-MS/MS identification. Highly abundant adulterant proteins from HWP derive from gluten proteins in wheat seeds (Garcia-Molina et al., 2017). In addition, the top 10 most abundant proteins from milk were also defined in our study (Table 3-S2). All these proteins were identified in adulterated milk

4. Conclusion

In our study, high speed centrifugation at 20 000 g for 60 min was found to be an effective pre-treatment to reduce highly abundant milk protein in milk samples before MS analysis. LC-MS/MS protein fingerprints coupled with PCA successfully differentiated adulterated samples (SP and HWP at level of 0.5-4%, PP and HRP at level of 2% and 4 %) from authentic milk, and subsequent protein identification

allowed the identification of adulterants (SP, PP, and HWP) used in milk adulteration. However, no rice protein was identified in the samples adulterated with HRP. The identification of adulterants protein by LC-MS/MS may be disturbed by the degree of hydrolysis of plant protein.

5. Acknowledgements

We gratefully acknowledge financial support of the Special Fund for Agro-scientific Research in the Public Interest (201403071), Modern Agro-Industry Technology Research System of China (CARS-37), and the Project of Risk Assessment on Raw Milk (GJFP2017008) from Chinese Ministry of Agriculture. We thank the University of Liège-Gembloux Agro-Bio Tech and the research platform AgricultureIsLife. Moreover, we would like to thank Prof. Georges Lognay (ULg-Belgium) for his suggestions for writing the paper.

6. References

Aaslyng, M. D., M. Martens, L. Poll, P. M. Nielsen, H. Flyge, and L. M. Larsen. 1998. Chemical and sensory characterization of hydrolyzed vegetable protein, a savory flavoring. J Agric Food Chem 46(2):481-489.

Cordewener, J. H. G., D. M. A. M. Luykx, R. Frankhuizen, M. G. E. G. Bremer, H. Hooijerink, and A. H. P. America. 2009. Untargeted LC-Q-TOF mass spectrometry method for the detection of adulterations in skimmed-milk powder. J Sep Sci 32(8):1216-1223.

Du, L., W. Lu, Z. Cai, L. Bao, C. Hartmann, B. Gao, and L. Yu. 2018. Rapid detection of milk adulteration using intact protein flow injection mass spectrometric fingerprints combined with chemometrics. Food Chem 240(Supplement C):573-578.

Fox, P. F., T. Uniacke-Lowe, P. L. H. McSweeney, and J. A. O'Mahony. 2015. Dairy Chemistry and Biochemistry. 2nd ed. Milk proteins. Springer International, Switzerland.

Garcia-Molina, M. D., V. Muccilli, R. Saletti, S. Foti, S. Masci, and F. Barro. 2017. Comparative proteomic analysis of two transgenic low-gliadin wheat lines and non-transgenic wheat control. J Proteomics 165:102-112.

Haasnoot, W. and J. G. du Pre. 2007. Luminex-based triplex immunoassay for the simultaneous detection of soy, pea, and soluble wheat proteins in milk powder. J Agr Food Chem 55(10):3771-3777.

IDF. 2014. milk products-Determination of nitrogen content-Part 1: Kjeldahl principle and crude protein calculation. IDF Standard:020-021.

Jablonski, J. E., J. C. Moore, and J. M. Harnly. 2014. Nontargeted detection of adulteration of skim milk powder with foreign proteins using UHPLC-UV. J Agric Food Chem 62(22):5198-5206.

Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227(5259):680-685.

Lopez-Tapia, J., M. R. Garcia-Risco, M. A. Manso, and R. Lopez-Fandino. 1999. Detection of the presence of soya protein in milk powder by sodium dodecyl sulfate

capillary electrophoresis. J Chromatogr A 836(1):153-160.
Lu, W., J. Liu, B. Gao, X. Lv, and L. L. Yu. 2017. Technical note: Nontargeted detection of adulterated plant proteins in raw milk by UPLC-quadrupole time-of-flight mass spectrometric proteomics combined with chemometrics. J Dairy Sci 100(9):6980-6986.
Luykx, D. M. A. M., J. H. G. Cordewener, P. Ferranti, R. Frankhuizen, M. G. E. G. Bremer, H. Hooijerink, and A. H. P. America. 2007. Identification of plant proteins in adulterated skimmed milk powder by high-performance liquid chromatography-mass spectrometry. J Chromatogr A 1164(1-2):189-197.
Manso, M. A., T. M. Cattaneo, S. Barzaghi, C. Olieman, and R. Lopez-Fandino. 2002. Determination of vegetal proteins in milk powder by sodium dodecyl sulfate-capillary gel electrophoresis: interlaboratory study. J Aoac Int 85(5):1090-1095.
Moore, J. C., J. Spink, and M. Lipp. 2012. Development and application of a database of food ingredient fraud and economically motivated adulteration from 1980 to 2010. J Food Sci 77(4):R118-126.
Nakamura, R. and R. Teshima. 2013. Proteomics-based allergen analysis in plants. J Proteomics 93:40-49.
Nascimento, C. F., P. M. Santos, E. R. Pereira-Filho, and F. R. P. Rocha. 2017. Recent advances on determination of milk adulterants. Food Chem 221:1232-1244.
Scholl, P. F., S. M. Farris, and M. M. Mossoba. 2014. Rapid Turbidimetric Detection of Milk Powder Adulteration with Plant Proteins. J Agr Food Chem 62(7):1498-1505.
Tessier, B., M. Schweizer, F. Fournier, X. Framboisier, I. Chevalot, R. Vanderesse, C. Harscoat, and I. Marc. 2005. Prediction of the amino acid composition of small peptides contained in a plant protein hydrolysate by LC–MS and CE–MS. Food Res Int 38(5):577-584.
Yang, Y. X., N. Zheng, J. H. Yang, D. P. Bu, J. Q. Wang, L. Ma, and P. Sun. 2014. Animal species milk identification by comparison of two-dimensional gel map profile and mass spectrometry approach. Int Dairy J 35(1):15-20.

7. Supporting information

Additional Supporting information may be found in the online version of this article:

Table 3-S1: Top 10 proteins identified from each adulterant in corresponding adulterated milk

Adulte-rants	Protein name	Organism	Protein numbers	Razor + unique peptides	Mass (kDa)	Coverage (%)	Score
Soy protein	Alpha' subunit of β-conglycinin	*Glycine max*	9	43	65.14	59.2	323.31
	Alpha subunit of β-conglycinin	*Glycine max*	10	33	63.16	63.5	323.31
	Glycinin A3B4 subunit	*Glycine soja*	13	19	57.70	70.6	323.31
	Glycinin	*Glycine max*	6	32	63.80	62.7	323.31
	Uncharacterized protein	*Glycine max*	5	7	18.46	43.7	221.18
	Kunitz trypsin inhibitor	*Glycine max*	23	13	24.14	44.9	311.72
	Uncharacterized protein	*Glycine max*	3	25	57.99	65.1	323.31
	Beta-conglycinin β subunit	*Glycine max*	1	24	48.33	67.1	323.31
	Seed maturation protein PM31	*Glycine max*	2	11	17.75	52.9	248.57
	Uncharacterized protein	*Glycine max*	1	27	54.68	72.5	323.31
Pea protein	Convicilin	*Pisum sativum*	1	34	72.06	65.9	323.31
	Vicilin	*Pisum sativum*	4	24	64.6	47.3	323.31
	Legumin A2	*Pisum sativum*	2	32	59.27	61.9	323.31
	Provicilin	*Pisum sativum*	2	12	31.54	70.5	323.31
	Vicilin 47k	*Pisum sativum*	1	3	49.66	72.1	255.19
	P54 protein	*Pisum sativum*	1	28	54.66	61.5	323.31
	Legumin (Minor small)	*Pisum sativum*	1	20	64.87	55.1	323.31

Table 3-S1 continued

Adulterants	Protein name	Organism	Protein numbers	Razor + unique peptides	Mass (kDa)	Coverage (%)	Score
Pea protein	LegA class	Pisum sativum	3	10	58.79	61.9	323.31
	Vicilin, 14 kDa component	Pisum sativum	1	2	14.04	51.6	89.15
	Legumin J	Pisum sativum	1	13	56.90	69.2	323.31
Hydrol-yzed wheat protein	Alpha-amylase inhibitor CM3	Triticum turgidum subsp. Durum	3	8	18.22	75.6	323.31
	0.19 dimeric α-amylase inhibitor	Triticum aestivum	25	5	13.34	91.1	323.31
	0.19 dimeric α-amylase inhibitor	Triticum aestivum	8	8	13.25	89.5	323.31
	Low molecular weight glutenin subunit	Triticum aestivum	95	3	34.78	18.5	323.31
	Gamma-gliadin	Triticum dicoccoides	1	2	14.6	29.2	94.17
	Alpha-amylase/trypsin inhibitor CM3	*Triticum aestivum*	1	2	17.30	52.5	19.22
	High molecular weight glutenin subunit Bx17	*Triticum aestivum*	43	6	80.07	12.6	100.22
	Uncharacterized protein	*Triticum aestivum*	10	5	21.89	35.8	123.43
	Alpha-amylase inhibitor CM1	*Triticum aestivum*	4	5	13.10	68.3	323.31
	Putative α-amylase inhibitor CM2	*Triticum aestivum*	1	4	13.01	60.0	321.18

Table 3-S2: Top 10 identified milk proteins in unadulterated (control) milk

Protein name	Organism	Number of proteins	Razor + unique peptides	Mass (kD)	Coverage (%)	Score
Beta-lactoglobulin	*Bos taurus*	7	24	19.97	86.5	323.31
Kappa-casein	*Bos indicus*	20	11	17.33	58.3	323.31
Alpha-S1-casein	*Bos taurus*	3	17	24.53	59.3	323.31
Serum albumin	*Bos taurus*	5	58	69.32	74.5	323.31
Alpha-lactalbumin	*Bos taurus*	7	18	14.16	65.9	323.31
Lactoferrin	*Bos taurus*	14	47	76.27	67.1	323.31
Lactadherin	*Bos taurus*	5	26	47.41	58.3	323.31
Beta-lactoglobulin	*Bos taurus*	2	4	19.88	86.5	323.31
Beta-casein	*Bos taurus*	2	5	25.00	25.0	187.74
Polymeric immunoglobulin receptor	*Bos taurus*	2	30	82.43	47.3	323.31

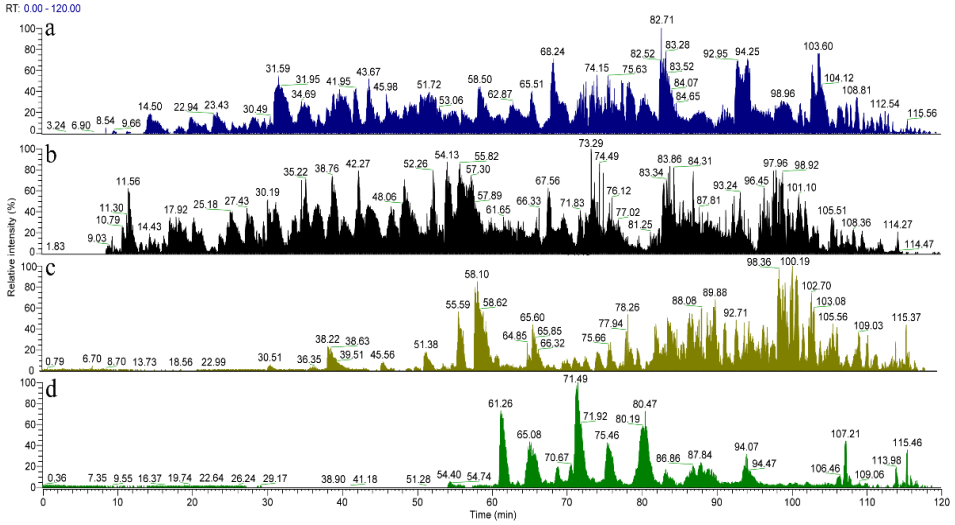


Figure 3-S1: Base peak chromatogram of mass spectra of soy protein solution (a), pea protein solution (b), hydrolyzed wheat protein solution (c), and hydrolyzed rice protein solution (d)

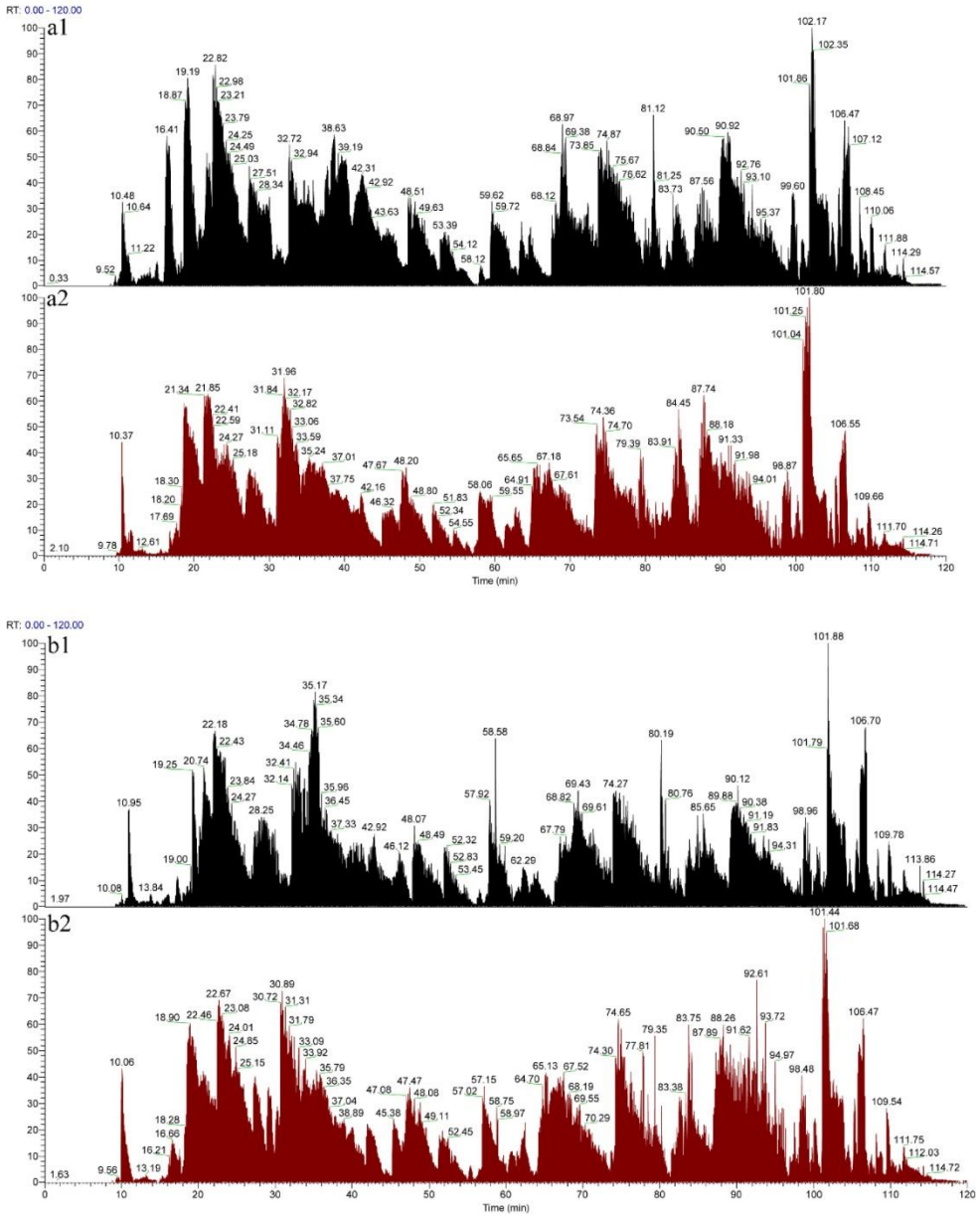


Figure 3-S2: Base peak chromatogram of mass spectra of milk (a1, a2) and adulterated samples (b1, b2) with soy protein at 4% of total protein

4

Research advances in milk production and detection by infrared spectroscopy

As one of detection methods, the infrared spectroscopy has been extensively applied to each link in the chain of milk production, from prediction of milk composition to assessment of milk quality. Many milk composition contents have been predicted by infrared spectroscopy, including fat, protein, lactose, and mineral elements. Reports on milk adulteration identified by near- and mid- infrared spectroscopy was also reviewed in this chapter.

This chapter was extracted from one published paper, and related parts with thesis topic, such as, introduction, data processing and model establishment, and application of infrared spectroscopy on milk composition and milk adulteration, were selected. The whole paper would be presented as appendix at the end of this thesis.

Extraction from Yang, J., N. Zheng, Y. Yang, Y. Zhang, and S. Li. 2016. Research advances in milk production and detection by infrared spectroscopy. Transactions of the Chinese Society of Agricultural Engineering, 32(17): 1-11 (*in Chinese*).

1. Introduction

Compared with the time-consuming and labor-intensive laboratory reference methods, spectroscopy methods are becoming more popular for their rapid non-destructive testing. Infrared spectra (IR) are produced by the absorption change of infrared radiation interacting with molecular groups. According to the wavelength region, the spectra can be classified into near-infrared (NIR, 14 000-4 000 cm^{-1}), mid-infrared (MIR, 4 000-400 cm^{-1}), and far-infrared (FIR, 400-50 cm^{-1}). NIR and MIR are often used to detect, characterise, and quantify chemical components. NIR is the result of overtone and combination bands associated with the fundamental vibrations of hydrogen-containing functional groups. The band signal is relatively weak and suitable for direct analysis of highly-absorbed or strongly-scattered samples without pretreatment (Arbuckle et al., 1996; Rodriguez-Saona and Allendorf, 2011). MIR is the absorption band that is caused by the fundamental vibration of specific functional groups and can be used to identify the structure of organic components. The fingerprint area contains various structural information such as fats and proteins, and the ratio of band intensity to functional group concentration can be used for quantitative analysis (Paré and Bélanger, 1997; Rodriguez-Saona and Allendorf, 2011). Fourier transform (FT) devices improve the analysis speed and accuracy of spectroscopic techniques by resolving overlapping spectral bands, reducing bandwidth, and increasing peak height (Markovich and Pidgeon, 1991). Attenuated total reflectance (ATR) technology improves the accuracy of FTIR data, because multiple reflectance in samples increases the spectral response when compared with single-reflection crystals (Rodriguez-Saona and Allendorf, 2011).

NIR is widely used for the quantification of components in liquid milk and milk powder (Wu et al., 2008; Aernouts et al., 2011a; Inácio et al., 2011; Huang et al., 2014), identification of adulterants (Borin et al., 2006; Balabin and Smirnov, 2011; Huang et al., 2015), quality inspection (Al-Qadiri et al., 2008; Kong et al., 2013; Yazdanpanah and Langrish, 2013), and can be used for real-time on-line monitoring of raw milk production (Lyndgaard et al., 2012; Melfsen et al., 2012a; Santos et al., 2013b). MIR can not only accurately determine the milk composition, but also predict the milk fatty acids (FAs), and protein components. Through algorithm optimisation, spectral data standardisation between different instruments could be achieved. It has been proposed that a large spectral database based on networks across regions could be established to improve farm management (Grelet et al., 2015). This part focuses on the literature on IR applications for milk composition prediction and quality inspection, to provide some directions for future studies using IR.

2. Data processing and model establishment

Spectra are affected by many factors, such as the complexity and specificity of the absorption spectrum of chemical bonds, sample particle scattering and molecular interactions, variance in environmental conditions, differences in equipment performance, and so on. Band selection and data preprocessing are required to reduce the differences in data collection and improve model reliability (Zou et al., 2010; De

Marchi et al., 2014). Common band selection methods include manual selection, multiple linear regression (MLR), successive projection algorithm (SPA), uninformative variable elimination (UVE), artificial neural networks (ANN), and genetic algorithms (GA). Data preprocessing methods mainly include scattering correction and its derivatives (Zou et al., 2010).

Cross-validation for the whole data set would overestimate the predictive power of an IR model, therefore a small additional test set is necessary for external validation. The number of samples in the calibration set should account for 50% or 75% of the total data (Bittante et al., 2014). Qualitative models classify samples according to absorption peaks based on pattern recognition methods such as correlation, distance, and discriminant analysis (Roggo et al., 2007). Commonly used evaluation parameters include the false positive rate, false negative rate, sensitivity, specificity, etc. (de Roos et al., 2007; Botelho et al., 2015). The quantitative model is based on the regression model derived from the relationship between spectral data and dependent variables in the calibration set and predicts the dependent variable using spectral data in the validation set. The root-mean-square error of prediction (RMSEP), calculated from the predicted value and the measured value, or the standard error of prediction (SEP) and determination coefficient (R^2) are used to evaluate the model performance (Zou et al., 2010). The ratio-performance deviation (RPD), range error ratio (RER), relative prediction error (RPE), and concordance correlation coefficient (CCC) are also important evaluation parameters (De Marchi et al., 2014). The qualitative analysis methods include the Mahalanobis distance, partial least squares discriminant analysis (PLS-DA), soft independent modelling of class analogy (SIMCA), and principal component analysis (PCA), etc., while the quantitative model commonly uses partial least square regression (PLSR), support vector machine (SVM), and ANN (Rodriguez-Saona and Allendorf, 2011; Domingo et al., 2014).

3. Determination of milk composition and quality

NIR predictions for milk composition are influenced by the spectral region, sample thickness, and measurement modes. The best accuracy for a NIR model was obtained with long wavelength bands (1 100 to 2 400 nm), 1 mm sample thickness, and the first derivative data transformation. For short wavelengths from 700 to 1 100 nm, the best accuracy for fat was obtained with a 10 mm sample, and for total protein with a 1 mm sample thickness. Lactose prediction was less affected by the sample thickness and spectral region (Tsenkova et al., 1999). NIR in reflectance mode resulted in accurate prediction of fat and crude protein in milk ($R^2 > 0.95$) and poor lactose prediction ($R^2 < 0.75$). In contrast, the transmittance spectra can achieve more accurate predictions for these items, and the corresponding R^2 of prediction were 0.99, 0.93, and 0.88 (Aernouts et al., 2011a). Moreover, some studies have shown that NIR (851-1 649 nm) in diffusion reflectance has a similar or better prediction for fat, protein, and lactose in milk, compared with transmittance or transflectance mode (Melfsen et al., 2012b). NIR in diffusion reflectance not only accurately predicted fat, protein, and lactose in milk (R^2 = 0.99, 0.98, and 0.92, respectively, SEP = 0.09, 0.05, and 0.06), but also achieved a good prediction for urea and somatic cell counts (logarithmic

transformation), R^2 of prediction was 0.82 and 0.85, and SEP was 19.3 mg/L and 0.18, respectively (Melfsen et al., 2012a). The presence of too many somatic cells in milk affected the NIR prediction for milk composition, so sorting of raw milk by somatic cell counts is necessary before dataset partition (Tsenkova et al., 2001). The prediction parameters for the main components in milk using NIR models from different studies are shown in Table 4-1, and NIR in diffuse reflectance is widely used. The good prediction performance of the NIR model makes it possible to assess fresh milk quality in real time, provide farmers with milk composition information and dairy cows' physiological status, and thereby to improve the efficiency of milk production (Kawasaki et al., 2008).

Table 4-1: Model performance of near-infrared spectroscopy for major milk components

Spectral mode	Fat		Protein		Lactose		References
	R^2	RMSEP	R^2	RMSEP	R^2	RMSEP	
Diffuse reflectance	0.977	0.154	0.960	0.134	-	-	Wang et al. (2015)
Transmittance	0.998	0.001	0.998	0.001	-	-	Zhao et al. (2014)
Fourier transform	0.995	0.136	0.975	0.195	-	-	Zhang (2010)
Transflectance	0.903	0.225	0.959	0.048	0.902	0.044	Yang et al. (2013)
Diffuse reflectance	0.998	0.09	0.98	0.05	0.92	0.06	Melfsen et al. (2012a)
Diffuse reflectance	0.95	0.25	0.83	0.26	0.72	0.15	Kawasaki et al. (2008)
Diffuse reflectance (Transmittance for lactose)	0.997	0.047	0.959	0.099	0.883	0.115	Aernouts et al. (2011a)

Pretreatment of milk samples, such as the addition of preservatives and homogenisation, would influence the MIR prediction. The addition of 0.02% potassium dichromate had little influence on the results of the MIR detection, whereas bromo-n-propylene glycol (0.02%) preserved milk had higher protein readings (a positive bias of about 0.01%) than potassium dichromate preserved or unpreserved milks. During cold preservation, uncorrected MIR readings for milk increased with the storage time, the growth rate was higher for raw milk than for pasteurised milk, and the stability of the instrument zero was lower for raw milk than for pasteurised milk (Barbano et al., 2010). The prediction results for ATR on milk composition are better than those of high-throughput transmission spectra. Homogenisation was crucial to obtain a good fat prediction, but had little effect on the prediction of other components (Aernouts et al., 2011b). IR can predict not only the milk composition, but these spectroscopies can also detect the protein fraction, FA composition, and other trace substances in milk.

3.1. Protein

Determination of milk proteins related to the characteristic absorption of amide I and II bands at 1 700 to 1 500 cm^{-1}, and phosphate groups bound to casein at 1 100 to 1 060 cm^{-1}, other milk components (fat and lactose), and protein particles would affect the prediction of the PLS model for milk proteins (Etzion et al., 2004). A suitable region selection algorithm, SIMPLe-to-use Interactive Self-modeling Mixture Analysis, combined with IR can quantitatively predict the secondary structure of the polypeptide chain, and the correlation coefficient of cross-validation between the predicted and measured values of the α-helix and β-sheet was 0.86-0.98 (Bogomolov and Hachey, 2007). MIR may not be ideal for predicting individual milk protein composition with high accuracy. The R^2 of cross-validation (R^2cv) of MIR prediction of casein (CN), αs1CN, αs2CN, βCN, κCN, and γCN (g/L milk) in milk were 0.77, 0.66, 0.49, 0.53, 0.63, and 0.60, respectively, while the R^2cv for whey protein, alpha lactalbumin, and beta lactoglobulin (g/L milk) were 0.61, 0.31, and 0.64, respectively (Bonfatti et al., 2011). Other studies have similar predictions of whey protein and its fractions (alpha lactalbumin, and beta lactoglobulin), and poor prediction of total and individual caseins, using raw MIR spectra (De Marchi et al., 2009; Rutten et al., 2011a), whereas there are some reports that predict the total CN with validation $R^2 > 0.90$ (Luginbuhl, 2002). The difference between these models may be associated to the reference methods for protein determination used in these studies (De Marchi et al., 2014). The MIR prediction for milk protein composition can be used to estimate breeding values and improve protein composition on a genetic level (Rutten et al., 2011a). FTIR combined with PLS can distinguish the milk produced by goats with two weak haploids from others, thereby selection of goats with high casein expression or screening for milk samples with high casein content is possible (Berget et al., 2010). FTIR prediction of β-LG genotypes showed a repeatability of 0.85, and it can improve the percentage of correctly predicted β-LG genotypes, in combination with pedigree information and derived genotypes (Rutten et al., 2011b).

3.2. Milk Fatty Acid Composition

There are two absorption bands for milk fat in the MIR region, fat A at 5.73 μm and fat B at 3.48 μm, which involved the stretch of C=O and C–H, respectively (Biggs and McKenna, 1989). Fat B MIR predictions increased and fat A MIR prediction decreased relative to reference chemistry with increasing FA chain length. When MIR fat prediction of fat B was corrected according to unsaturation variation between samples, fat B had a positive correlation with the FA chain length (correlation coefficient was 0.42-0.89); when the corrected ratio of fat B for unsaturation was 45:55, fat A gave the best fit between MIR prediction and the reference (Kaylegian et al., 2009). Oleic acid and linoleic acid presented different spectra in the MIR region. Oleic acid has two characteristic peaks at 1 119 and 1 091 cm^{-1}, while the characteristic peaks of linoleic acid appear at 1 048, 1102, and 1 121 cm^{-1} (Yang, 2011). The IR prediction of milk FAs in different studies are shown in Tables 4-2, 4- 3, and 4-4. Among unsaturated FAs, a better prediction accuracy for c9C18:1 is observed; meanwhile, the prediction accuracy for saturated FAs and monounsaturated

FAs is greater than that of polyunsaturated FAs, which may be in line with individual and grouped FA concentration in milk (Soyeurt et al., 2006; Rutten et al., 2009; De Marchi et al., 2011). There are two ways of expressing FAs in milk, namely the concentration of FAs in milk (g/L milk or g/kg milk) and the FA content in fat (g/kg total FAs). The accuracy of MIR predictions expressed as the FA concentration are better than those of models expressed as FA content (Soyeurt et al., 2006; De Marchi et al., 2014), which is similar to the NIR prediction for liquid milk (Coppa et al., 2014). However, the comparison of prediction accuracy between MIR and NIR for FAs in oven-dried milk varied with FAs and expression. When expressed as g/kg of milk, the accuracy of NIR prediction was worse than MIR for almost all FAs. When expressed as g/100 g total FAs, MIR and NIR shared a similar prediction accuracy for the group of even-chain saturated FA, odd-chain FA, unsaturated FA, conjugated linoleic acid, n-3 FA, and c9C18:1/C16 ratio; while monounsaturated FA, n-6/n-3 ratio, polyunsaturated FA (PUFA), and n-6 FA were better predicted by NIR (Coppa et al., 2014). High levels of FAs in milk, such as even-chain FAs, could achieve good MIR prediction fitting with measured values, no matter the expression of FAs (Soyeurt et al., 2006). Compared with predictions for liquid milk, NIR quantification of milk FA was more accurate or similar for oven-dried milk (Coppa et al., 2010; Coppa et al., 2014), but the reliability decreased for thawed liquid milk (Coppa et al., 2014). MIR prediction was also used to estimate heritability and correlation of FAs in goat milk (Maroteau et al., 2014). Strong relationships between the sample size of calibration and validation R^2, as well as strong genetic correlations were observed. As the calibration number increased, the variation range of the validation R^2 and the genetic correlation coefficient gradually narrowed. When there were 1 000 samples in calibration, the genetic correlation changed within a range of 0.1 (Rutten et al., 2010).

Table 4-2: Model prediction of infrared spectroscopy for milk fatty acids (saturated fatty acid)

Sample size	Unit	C4:0	C6:0	C8:0	C10:0	C12:0	C14:0	C15:0	C16:0	C17:0	C18:0	References
267	g/L of cow milk	-	-	0.74 (0.07)	0.73 (0.19)	0.75 (0.25)	0.77 (0.6)	0.63 (0.07)	0.70 (1.59)	0.56 (0.03)	0.65 (0.75)	De Marchi et al. (2011) [1]
600	g/L of cow milk	0.51 (0.08)	0.52 (0.04)	0.59 (0.02)	0.64 (0.04)	0.74 (0.02)	0.82 (0.05)	0.40 (0.01)	0.82 (0.17)	-	0.69 (0.13)	Soyeurt et al. (2006) [2]
600	g/kg of milk fat	0.39 (1.60)	0.41 (0.98)	0.46 (0.50)	0.53 (0.90)	0.64 (0.53)	0.67 (1.14)	0.53 (0.2)	0.50 (3.5)	-	0.09 (2.77)	
517	g/L of cow milk	0.94 (0.01)	0.97 (0.00)	0.97 (0.00)	0.96 (0.01)	0.96 (0.01)	0.97 (0.02)	-	0.95 (0.08)	0.89 (0.00)	0.90 (0.05)	Soyeurt et al. (2011) [2]
3 660	g/L of cow milk	0.91 (0.10)	0.96 (0.20)	0.94 (0.50)	0.92 (0.10)	0.85 (0.30)	0.94 (0.3)	-	0.94 (0.10)	-	0.82 (0.70)	Rutten et al. (2009) [3]
3 660	g/kg of milk fat	0.55 (0.00)	0.73 (0.30)	0.73 (0.60)	0.75 (0.20)	0.68 (0.30)	0.73 (0.3)	-	0.71 (0.00)	-	0.51(1.20)	
238-241[a]	g/L of cow milk	0.93 (0.006)	0.96 (0.003)	0.96 (0.002)	0.95 (0.007)	0.95 (0.008)	0.94 (0.024)	-	0.94 (0.066)	-	0.84 (0.041)	
98-104[b]	g/L of cow milk	0.61 (0.01)	0.86 (0.004)	0.89 (0.003)	0.85 (0.011)	0.82 (0.018)	0.84 (0.03)	-	0.82 (0.111)	-	0.49 (0.054)	Ferrand-Calmels et al. (2014) [4]
135-140	g/L of sheep milk	0.93 (0.01)	0.97 (0.005)	0.96 (0.008)	0.93 (0.041)	0.97 (0.019)	0.96 (0.045)	-	0.94 (0.091)	-	0.83 (0.061)	
215-229	g/L of goat milk	0.96 (0.004)	0.95 (0.004)	0.97 (0.004)	0.98 (0.013)	0.92 (0.013)	0.93 (0.023)	-	0.96 (0.042)	-	0.86 (0.034)	
154	g/L of cow milk	0.87 (0.009)	0.97 (0.003)	0.97 (0.002)	0.95 (0.008)	0.95 (0.011)	0.83 (0.040)	0,67 (0.005)	0.91 (0.087)	0.74 (0.002)	0.75 (0.048)	Ferrand et al. (2011) [2]

Table 4-2 continued

Sample size	Unit	C4:0	C6:0	C8:0	C10:0	C12:0	C14:0	C15:0	C16:0	C17:0	C18:0	References
1 167-1 187	g/L of cow milk	0.93 (0.008)	0.96 (0.005)	0.96 (0.003)	0.96 (0.008)	0.95 (0.01)	0.95 (0.028)	-	0.97 (0.068)	0.89 (0.003)	0.90 (0.045)	Maurice-Van Eijndhoven et al. (2013) [2]
279-344	g/kg of total FA	0.66 (0.42)	0.88 (0.21)	0.90 (0.13)	0.91 (0.34)	0.89 (0.41)	0.88 (1.07)	0.53 (0.14)	0.91 (2.20)	0.65 (0.08)	0.80 (1.31)	Coppa et al. (2010) [5]

Parameters of models in this table are as follows: 1) correlation coefficient of cross-validation (RMSE), 2) determination coefficient of cross-validation (SE), 3) determination coefficient of validation (bias/mean × 100), 4) determination coefficient of validation (residual error), 5) determination coefficient of validation (prediction error); 1, 2, 3, 4 used the mid-infrared spectroscopy model; 5 used the near-infrared spectroscopy model, spectra scanned for oven dried milk

a) data from MilkoScan FT6000, b) data from Bentley FTS

Table 4-3: Model prediction of infrared spectroscopy for milk fatty acids (unsaturated fatty acids and fatty acid groups)

Sample size	Unit	c9C14:1	c9C16:1	c9C18:1	c11C18:1	c9c12C18:2	C18:3n-3	c9t11C18:2	References
267	g/L of cow milk	0.68 (0.08)	0.60 (0.11)	0.73 (1.13)	0.59 (0.04)	-	0.51 (0.04)	0.58 (0.04)	De Marchi et al. (2011)[1]
600	g/L of cow milk	0.07 (0.01)	0.65 (0.02)	-	-	0.62 (0.02)	0.14 (0.01)	0.07 (0.02)	(Soyeurt et al., 2006) [2]
600	g/kg of milk fat	0.23 (0.28)	0.37 (0.37)	-	-	0.11 (0.44)	0.20 (0.20)	0.34 (0.37)	
517	g/L of cow milk	0.68 (0.01)	0.71 (0.01)	0.97 (0.05)	-	0.74 (0.01)	0.71 (0.01)	0.74 (0.01)	Soyeurt et al. (2011) [2]
3 660	g/L of cow milk	-	-	0.92 (0.30)	0.27 (0.10)	0.36 (0.90)	0.45 (3.30)	0.58 (1.00)	Rutten et al. (2009) [3]
3 660	g/kg of milk fat	-	-	0.84 (0.50)	0.22 (0.40)	0.28 (0.60)	0.38 (2.80)	0.56 (1.10)	
238-241[a]	g/L of cow milk	-	-	0.96 (0.039)	-	0.77 (0.006)	0.85 (0.004)	0.82 (0.003)	
98-104[b]	g/L of cow milk	-	-	0.86 (0.063)	-	0.75 (0.006)	0.81 (0.003)	0.64 (0.003)	Ferrand-Calmels et al. (2014) [4]
135-140	g/L of cow milk	-	-	0.97 (0.057)	-	0.49 (0.012)	0.74 (0.007)	0.91 (0.011)	
215-229	g/L of cow milk	-	-	0.95 (0.037)	-	0.89 (0.007)	0.79 (0.003)	0.71 (0.003)	
154	g/L of cow milk	-	-	-	-	0.76 (0.006)	0.85 (0.003)	0.66 (0.004)	Ferrand et al. (2011) [2]

Table 4-3 continued

Sample size	Unit	c9C14:1	c9C16:1	c9C18:1	c11C18:1	c9c12C18:2	C18:3n-3	c9t11C18:2	References
1 167-1 187	g/L of cow milk	0.78 (0.007)	0.78 (0.011)	-	-	-	-	-	Maurice-Van Eijndhoven et al. (2013) [3]
279-344	g/kg of total FA	0.57 (0.22)	0.44 (0.25)	0.93 (1.77)	0.29 (0.13)	0.34 (0.28)	0.48 (0.16)	0.73 (0.87)	Coppa et al. (2010) [5]

Parameters of models in this table are as follows: 1) correlation coefficient of cross-validation (RMSE), 2) determination coefficient of cross-validation (SE), 3) determination coefficient of validation (bias/mean × 100), 4) determination coefficient of validation (residual error), 5) determination coefficient of validation (prediction error); 1, 2, 3, 4 used the mid-infrared spectroscopy model; 5 used the near-infrared spectroscopy model, spectra scanned for oven dried milk

a) data from MilkoScan FT6000, b) data from Bentley FTS

Table 4-4: Model prediction of infrared spectroscopy for milk fatty acids (fatty acid groups)

Sample size	Unit	SCFA	MCFA	LCFA	SFA	MUFA	PUFA	UFA	References
267	g/L of cow milk	-	0.73 (2.66)	0.76 (1.94)	0.72 (2.97)	0.74 (1.39)	0.64 (0.22)	0.71 (1.57)	De Marchi et al. (2011)[1]
600	g/L of cow milk	-	-	-	0.94 (0.20)	0.85 (0.22)	0.39 (0.04)	0.66 (0.34)	Soyeurt et al. (2006) [2]
600	g/kg of milk fat	-	-	-	0.63 (3.75)	0.52 (4.10)	0.10 (0.74)	0.63 (3.75)	
517	g/L of cow milk	0.98 (0.02)	0.98 (0.09)	0.98 (0.09)	1.00 (0.05)	0.99 (0.04)	0.85 (0.02)	0.99 (0.04)	Soyeurt et al. (2011) [2]
3 660	g/L of cow milk	0.95 (0.00)	0.97 (0.00)	-	-	-	-	-	Rutten et al. (2009) [3]
3 660	g/kg of milk fat	0.82 (0.30)	0.77 (0.10)	-	-	-	-	-	
238-241[a]	g/L of cow milk	-	-	-	1.00 (0.035)	0.97 (0.037)	0.76 (0.01)	0.98 (0.038)	Ferrand-Calmels et al. (2014) [4]
98-104[b]	g/L of cow milk	-	-	-	0.96 (0.09)	0.89 (0.068)	0.60 (0.01)	0.83 (0.10)	
135-140	g/L of cow milk	-	-	-	1.00 (0.049)	0.99 (0.044)	0.96 (0.015)	0.99 (0.048)	
215-229	g/L of cow milk	-	-	-	0.99 (0.043)	0.96 (0.037)	0.92 (0.01)	0.97 (0.039)	
154	g/L of cow milk	-	-	-	0.99 (0.045)	0.97 (0.044)	0.62 (0.010)	-	Ferrand et al. (2011) [2]

Table 4-4 continued

Sample size	Unit	SCFA	MCFA	LCFA	SFA	MUFA	PUFA	UFA	References
1167-1 187	g/L of cow milk	0.96 (0.02)	0.98 (0.086)	-	1.00 (0.051)	-	-	-	Maurice-Van Eijndhoven et al. (2013) [3]
279-344	g/kg of total FA	-	-	-	0.97 (1.94)	0.97 (1.81)	0.85 (0.87)	0.97 (2.23)	Coppa et al. (2010) [5]

Parameters of models in this table are as follows: 1) correlation coefficient of cross-validation (RMSE), 2) determination coefficient of cross-validation (SE), 3) determination coefficient of validation (bias/mean×100), 4) determination coefficient of validation (residual error), 5) determination coefficient of validation (prediction error); 1, 2, 3, 4 used the mid-infrared spectroscopy model; 5 used the near-infrared spectroscopy model, spectra scanned for oven dried milk

a) data from MilkoScan FT6000, b) data from Bentley FTS

SCFA: short chain fatty acids; MCFA: medium chain fatty acids; LCFA: long chain fatty acids; SFA: saturated fatty acid; MUFA: monounsaturated fatty acids; PUFA: polyunsaturated fatty acids; UFA: unsaturated fatty acids.

3.3. Other components

MIR can be used to predict cheese production. Favourable predictions were observed for the yield of total solid cheese and fresh cheese, with R^2cv of 0.95 and 0.83, respectively, and promising results were obtained for the recovered protein, total solids, and energy (R^2cv were 0.81, 0.86, and 0.76, respectively) (Ferragina et al., 2013). Visible and short wavelength NIR diffuse reflectance spectroscopy (600-1 000 nm) can be used to monitor spoilage of pasteurised skimmed milk by predicting the bacterial counts and pH of milk (R^2cv = 0.99 and 0.99, SEP = 0.34 cfu/mL and 0.031, respectively) (Al-Qadiri et al., 2008). FT-MIR can also accurately determine the titration acidity of milk, with R^2cv of 0.96, RMSE of the cross validation set of 0.72°T, and the RPD of 5.1 (Calamari et al., 2016). The pretreated NIR spectra combined multivariate regression model could predict the IgG level in colostrum, where the R^2 of the calibration set and cross validation set were 0.95 and 0.94, respectively (Rivero et al., 2012). The potential for MIR prediction of lactoferrin was also confirmed by studies, with R^2 of cross validation and external validation of 0.71 and 0.60 (Soyeurt et al., 2012), and there was a positive correlation between predicted lactoferrin and somatic cell scores, but a negative genetic correlation between predicted lactoferrin and milk yield was also observed (Soyeurt et al., 2007). When MIR combined with atomic absorption spectrometry was used to predict major mineral elements Ca, K, Mg, Na, and P in milk, only Ca, Na, and P showed the sufficient R^2cv (0.80, 0.70, and 0.79) for potential application. Finally, potential application of Ca and P equations were confirmed, whose R^2 of external validation were 0.97 and 0.88 (Soyeurt et al., 2009). Accurate MIR prediction for Ca and P was confirmed by other studies, and the contents of these elements were closely related to the agglutination traits of milk (Toffanin et al., 2015). FT-MIR can predict tetracycline levels in milk, validation R^2 reached 0.85-0.89, the detection range was 4-2 000 μg/kg, and SEP was from 89 to 387 μg/kg (Sivakesava and Irudayaraj, 2002).

4. Milk quality inspection

Combined with SIMCA, short-wave NIR can distinguish milk stored for 30 hours at different temperatures (6, 21, and 37°C) from control samples, with an accuracy of about 90% (Al-Qadiri et al., 2008). FT-NIR combined with principal component analysis can accurately recognise different brands of milk with an accuracy of 100% (Jin et al., 2016). And coupled with Fisher's multi-class linear discriminant analysis, FT-NIR can identify milk adulteration with plant cream, vegetable protein, and starch, and the correct rate was achieved in more than 94% of cases (Li and Ding, 2010). Compared with 1 704-1 400 cm^{-1} in the MIR region, 4 800-4 200 cm^{-1} in NIR was more sensitive to urea adulterated milk; for unknown samples, the prediction R^2 reached 0.999, and RMSEP was 0.219 g/L (Yang et al., 2012). Based on changes in fat and protein contents in milk within 52 hours, NIR could indicate changes in milk quality (Shi, 2014). Based on the prediction of pH and acidity of goat milk, NIR can also be used to evaluate the freshness of goat milk (Chu, 2012). Since phenolic compounds were stable during fermentation and manufacturing processing, it was

possible to discriminate probiotic milk samples according to the type of extract added and to evaluate the 'stability' of the product using NIR spectra combined with multivariate analysis (Aliakbarian et al., 2015). The adulterant levels in milk affected the discrimination of the NIR models, and the combination of non-linear pattern recognition and NIR could be useful for the identification and authentication of raw cow milks (Zhang et al., 2014). Nevertheless, some results indicated that the MIR system was superior to the NIR system in monitoring milk adulteration for additives such as water, whey protein, synthetic milk, synthetic urea, urea, and hydrogen peroxide (Santos et al., 2013b). A comprehensive index Q was constructed using milk indexes detected by a FTIR analyser, total solid-fat, ice-lactose, and lactose parameters were mainly included, and the addition of butter (> 0.058 g/100 g), gelatin hydrolysate (> 0.020 g/100 g), ammonium chloride (> 0.395 g/100 g), melamine (> 0.310 g/100 g), urea (> 0.443 g/100 g), sucrose and maltodextrin (> 0.024 g/100 g), whey (> 0.072 g/100 g), and milk powder and water (> 0.500 g/100 g) to milk could be recognised (Liu et al., 2015). FTIR could quantitatively detect the spiked level of baking soda, sodium citrate, and lactalbumin in milk, with all calibration R^2 above 0.91 and the detection limits of 0.015%, 0.017%, and 3.9%, respectively (Cassoli et al., 2011). ATR-MIR combined with PLS-DA was able to detect the presence of water, starch, sodium citrate, formaldehyde, and sucrose in milk, and the detection range was 0.5% to 10.0% (w/v) (Botelho et al., 2015). ATR-MIR also detected whey protein, hydrogen peroxide, synthetic urine, urea, and synthetic milk adulterated milk, and SEP was 2.33, 0.06, 0.41, 0.30, and 0.014 g/L, respectively, and the detection limits were 7.5, 0.019, 0.78, 0.78, and 0.1 g/L respectively (Santos et al., 2013a). Combined with PLS, single-beam ATR-FTIR can quickly predict melamine content in milk with the limits of detection and quantitation 2.5 and 15 mg/kg, respectively (Jawaid et al., 2013). If the correct data processing and multivariate algorithm was applied in the developed model, the detection limit of melamine for IR prediction could be less than 1 mg/kg, and a non-linear relationship was found between melamine content and IR response (Balabin and Smirnov, 2011; Domingo et al., 2014).

MIR was also sensitive to adulteration between different milks. The ratio of cow's milk, goat's milk, and sheep's milk in their mixture could be quantitatively predicted by FTIR. The prediction R^2 for a binary mixture was 0.91-0.98, RMSEP was 3.95%-8.03%; the prediction R^2 for a ternary mixture was 0.92-0.97, RMSEP ranged from 3.36% to 6.40% (Nicolaou et al., 2010). A main peak located at 1745 cm^{-1}, related to the degree of sugar carboxyl methyl esterification, was observed on the FTIR comparison between goat's and sheep's milk, and corresponding hierarchical and discriminant analyses showed goat samples could be separated from sheep samples (Pappas et al., 2008). Differences between soy milk and cow-buffalo milk, as well as their mixture at different ratios were centred on the MIR region at 1 680 to 1 058 cm^{-1}. PCA indicated that the addition above 5% of soybean milk to milk showed a significant difference from control milk. Based on the absorption of 1 472 to 1 241 cm^{-1}, the multivariate linear regression analysis showed that validation R^2 was 0.92 and the SEP was 7.56 for soybean milk levels in the mixture (Jaiswal et al., 2015).

5. References

Aernouts, B., E. Polshin, J. Lammertyn, and W. Saeys. 2011a. Visible and near-infrared spectroscopic analysis of raw milk for cow health monitoring: Reflectance or transmittance? Journal of Dairy Science 94(11):5315-5329.

Aernouts, B., E. Polshin, W. Saeys, and J. Lammertyn. 2011b. Mid-infrared spectrometry of milk for dairy metabolomics: A comparison of two sampling techniques and effect of homogenization. Anal Chim Acta 705(1-2):88-97.

Al-Qadiri, H. M., M. Lin, M. A. Al-Holy, A. G. Cavinato, and B. A. Rasco. 2008. Monitoring Quality Loss of Pasteurized Skim Milk Using Visible and Short Wavelength Near-Infrared Spectroscopy and Multivariate Analysis. Journal of Dairy Science 91(3):950-958.

Aliakbarian, B., M. Casale, M. Paini, A. A. Casazza, S. Lanteri, and P. Perego. 2015. Production of a novel fermented milk fortified with natural antioxidants and its analysis by NIR spectroscopy. LWT - Food Science and Technology 62(1):376-383.

Arbuckle, G. A., N. M. Buecheler, J. W. Hall, K. G. Valentine, S. Lefrant, J. Y. Mevellec, and E. Mulazzi. 1996. Spectroscopic properties of polyacetylenes synthesized via three modifications of Ziegler-Natta catalytic system. Synthetic Met 79(3):183-188.

Balabin, R. M. and S. V. Smirnov. 2011. Melamine detection by mid- and near-infrared (MIR/NIR) spectroscopy: A quick and sensitive method for dairy products analysis including liquid milk, infant formula, and milk powder. Talanta 85(1):562-568.

Barbano, D. M., K. L. Wojciechowski, and J. M. Lynch. 2010. Effect of preservatives on the accuracy of mid-infrared milk component testing. Journal of Dairy Science 93(12):6000-6011.

Berget, I., H. Martens, A. Kohler, S. K. Sjurseth, N. K. Afseth, B. Narum, T. Adnoy, and S. Lien. 2010. Caprine CSN1S1 haplotype effect on gene expression and milk composition measured by Fourier transform infrared spectroscopy. Journal of Dairy Science 93(9):4340-4350.

Biggs, D. A. and D. McKenna. 1989. Alternative methods for infrared analysis of fat in milk: interlaboratory study. Journal - Association of Official Analytical Chemists 72(5):724-734.

Bittante, G., A. Ferragina, C. Cipolat-Gotet, and A. Cecchinato. 2014. Comparison between genetic parameters of cheese yield and nutrient recovery or whey loss traits measured from individual model cheese-making methods or predicted from unprocessed bovine milk samples using Fourier-transform infrared spectroscopy. Journal Of Dairy Science 97(10):6560-6572.

Bogomolov, A. and M. Hachey. 2007. Application of SIMPLISMA purity function for variable selection in multivariate regression analysis: A case study of protein secondary structure determination from infrared spectra. Chemometr Intell Lab 88(1):132-142.

Bonfatti, V., G. Di Martino, and P. Carnier. 2011. Effectiveness of mid-infrared

spectroscopy for the prediction of detailed protein composition and contents of protein genetic variants of individual milk of Simmental cows. Journal Of Dairy Science 94(12):5776-5785.
Borin, A., M. F. Ferrao, C. Mello, D. A. Maretto, and R. J. Poppi. 2006. Least-squares support vector machines and near infrared spectroscopy for quantification of common adulterants in powdered milk. Anal Chim Acta 579(1):25-32.
Botelho, B. G., N. Reis, L. S. Oliveira, and M. M. Sena. 2015. Development and analytical validation of a screening method for simultaneous detection of five adulterants in raw milk using mid-infrared spectroscopy and PLS-DA. Food Chem 181:31-37.
Calamari, L., L. Gobbi, and P. Bani. 2016. Improving the prediction ability of FT-MIR spectroscopy to assess titratable acidity in cow's milk. Food Chem 192:477-484.
Cassoli, L. D., B. Sartori, and P. F. Machado. 2011. The use of the Fourier Transform Infrared spectroscopy to determine adulterants in raw milk. Revista Brasileira De Zootecnia-Brazilian Journal of Animal Science 40(11):2591-2596.
Chu, Y. 2012. Study on Discrimination of Fresh of Goat's Milk and Adulteration Goat's Milk by Near Infrared Spectroscopy Technology. Northwest A&F University, Yangling.
Coppa, M., A. Ferlay, C. Leroux, M. Jestin, Y. Chilliard, B. Martin, and D. Andueza. 2010. Prediction of milk fatty acid composition by near infrared reflectance spectroscopy. Int Dairy J 20(3):182-189.
Coppa, M., A. Revello-Chion, D. Giaccone, A. Ferlay, E. Tabacco, and G. Borreani. 2014. Comparison of near and medium infrared spectroscopy to predict fatty acid composition on fresh and thawed milk. Food Chem 150:49-56.
De Marchi, M., V. Bonfatti, A. Cecchinato, G. Di Martino, and P. Carnier. 2009. Prediction of protein composition of individual cow milk using mid-infrared spectroscopy. Ital J Anim Sci 8:399-401.
De Marchi, M., M. Penasa, A. Cecchinato, M. Mele, P. Secchiari, and G. Bittante. 2011. Effectiveness of mid-infrared spectroscopy to predict fatty acid composition of Brown Swiss bovine milk. Animal 5(10):1653-1658.
De Marchi, M., V. Toffanin, M. Cassandro, and M. Penasa. 2014. Invited review: Mid-infrared spectroscopy as phenotyping tool for milk traits. Journal of Dairy Science 97(3):1171-1186.
de Roos, A. P. W., H. J. C. M. van den Bijgaart, J. Hørlyk, and G. de Jong. 2007. Screening for Subclinical Ketosis in Dairy Cattle by Fourier Transform Infrared Spectrometry. Journal of Dairy Science 90(4):1761-1766.
Domingo, E., A. A. Tirelli, C. A. Nunes, M. C. Guerreiro, and S. M. Pinto. 2014. Melamine detection in milk using vibrational spectroscopy and chemometrics analysis: A review. Food Res Int 60:131-139.
Etzion, Y., R. Linker, U. Cogan, and I. Shmulevich. 2004. Determination of Protein Concentration in Raw Milk by Mid-Infrared Fourier Transform Infrared/Attenuated Total Reflectance Spectroscopy. Journal of Dairy Science 87(9):2779-2788.

Ferragina, A., C. Cipolat-Gotet, A. Cecchinato, and G. Bittante. 2013. The use of Fourier-transform infrared spectroscopy to predict cheese yield and nutrient recovery or whey loss traits from unprocessed bovine milk samples. Journal of Dairy Science 96(12):7980-7990.

Ferrand-Calmels, M., I. Palhière, M. Brochard, O. Leray, J. M. Astruc, M. R. Aurel, S. Barbey, F. Bouvier, P. Brunschwig, H. Caillat, M. Douguet, F. Faucon-Lahalle, M. Gelé, G. Thomas, J. M. Trommenschlager, and H. Larroque. 2014. Prediction of fatty acid profiles in cow, ewe, and goat milk by mid-infrared spectrometry. Journal of Dairy Science 97(1):17-35.

Ferrand, M., B. Huquet, S. Barbey, F. Barillet, F. Faucon, H. Larroque, O. Leray, J. M. Trommenschlager, and M. Brochard. 2011. Determination of fatty acid profile in cow's milk using mid-infrared spectrometry: Interest of applying a variable selection by genetic algorithms before a PLS regression. Chemometr Intell Lab 106(2):183-189.

Grelet, C., J. A. Fernández Pierna, P. Dardenne, V. Baeten, and F. Dehareng. 2015. Standardization of milk mid-infrared spectra from a European dairy network. Journal of Dairy Science 98(4):2150-2160.

Huang, Y., S. Min, J. Duan, L. Wu, and Q. Li. 2014. Identification of additive components in powdered milk by NIR imaging methods. Food Chem 145(0):278-283.

Huang, Y., K. Tian, S. Min, Y. Xiong, and G. Du. 2015. Distribution assessment and quantification of counterfeit melamine in powdered milk by NIR imaging methods. Food Chem 177(0):174-181.

Inácio, M. R. C., M. d. F. V. de Moura, and K. M. G. de Lima. 2011. Classification and determination of total protein in milk powder using near infrared reflectance spectrometry and the successive projections algorithm for variable selection. Vibrational Spectroscopy 57(2):342-345.

Jaiswal, P., S. N. Jha, A. Borah, A. Gautam, M. K. Grewal, and G. Jindal. 2015. Detection and quantification of soymilk in cow–buffalo milk using Attenuated Total Reflectance Fourier Transform Infrared spectroscopy (ATR–FTIR). Food Chem 168(0):41-47.

Jawaid, S., F. N. Talpur, S. T. H. Sherazi, S. M. Nizamani, and A. A. Khaskheli. 2013. Rapid detection of melamine adulteration in dairy milk by SB-ATR-Fourier transform infrared spectroscopy. Food Chem 141(3):3066-3071.

Jin, Y., B. Du, and X. Zhi. 2016. Fast recognition of adulterated milk and the brands of milk based on the NIR technology. Food Research And Development 37(3):178-181.

Kawasaki, M., S. Kawamura, M. Tsukahara, S. Morita, M. Komiya, and M. Natsuga. 2008. Near-infrared spectroscopic sensing system for on-line milk quality assessment in a milking robot. Computers and Electronics in Agriculture 63(1):22-27.

Kaylegian, K. E., D. A. Dwyer, J. M. Lynch, D. E. Bauman, J. R. Fleming, and D. M. Barbano. 2009. Impact of fatty acid composition on the accuracy of mid-infrared fat analysis of farm milks. Journal of Dairy Science 92(6):2502-2513.

Kong, W. W., C. Zhang, F. Liu, A. P. Gong, and Y. He. 2013. Irradiation dose

detection of irradiated milk powder using visible and near-infrared spectroscopy and chemometrics. Journal of Dairy Science 96(8):4921-4927.
Li, L. and W. Ding. 2010. Discriminant analysis of raw milk adulterated with botanical filling material using near infrared spectroscopy. Spectroscopy and Spectral Analysis 30(5):1238-1242.
Liu, J., J. Ren, Z. M. Liu, and B. H. Guo. 2015. A new comprehensive index for discriminating adulteration in bovine raw milk. Food Chem 172:251-256.
Luginbuhl, W. 2002. Evaluation of designed calibration samples for casein calibration in Fourier transform infrared analysis of milk. Lebensm-Wiss Technol 35(6):554-558.
Lyndgaard, C. B., S. B. Engelsen, and F. W. J. van den Berg. 2012. Real-time modeling of milk coagulation using in-line near infrared spectroscopy. Journal of Food Engineering 108(2):345-352.
Markovich, R. J. and C. Pidgeon. 1991. Introduction to Fourier transform infrared spectroscopy and applications in the pharmaceutical sciences. Pharmaceutical research 8(6):663-675.
Maroteau, C., I. Palhiere, H. Larroque, V. Clement, M. Ferrand, G. Tosser-Klopp, and R. Rupp. 2014. Genetic parameter estimation for major milk fatty acids in Alpine and Saanen primiparous goats. Journal Of Dairy Science 97(5):3142-3155.
Maurice-Van Eijndhoven, M. H., H. Bovenhuis, H. Soyeurt, and M. P. Calus. 2013. Differences in milk fat composition predicted by mid-infrared spectrometry among dairy cattle breeds in the Netherlands. Journal of Dairy Science 96(4):2570-2582.
Melfsen, A., E. Hartung, and A. Haeussermann. 2012a. Accuracy of in-line milk composition analysis with diffuse reflectance near-infrared spectroscopy. Journal of Dairy Science 95(11):6465-6476.
Melfsen, A., E. Hartung, and A. Haeussermann. 2012b. Accuracy of milk composition analysis with near infrared spectroscopy in diffuse reflection mode. Biosystems Engineering 112(3):210-217.
Nicolaou, N., Y. Xu, and R. Goodacre. 2010. Fourier transform infrared spectroscopy and multivariate analysis for the detection and quantification of different milk species. Journal of Dairy Science 93(12):5651-5660.
Pappas, C. S., P. A. Tarantilis, E. Moschopoulou, G. Moatsou, I. Kandarakis, and M. G. Polissiou. 2008. Identification and differentiation of goat and sheep milk based on diffuse reflectance infrared Fourier transform spectroscopy (DRIFTS) using cluster analysis. Food Chem 106(3):1271-1277.
Paré, J. R. J. and J. M. R. Bélanger. 1997. Instrumental methods in food analysis. Techniques and instrumentation in analytical chemistry. No. 18. Elsevier, Amsterdam ; New York.
Rivero, M. J., X. Valderrama, D. Haines, and D. Alomar. 2012. Prediction of immunoglobulin G content in bovine colostrum by near-infrared spectroscopy. Journal of Dairy Science 95(3):1410-1418.
Rodriguez-Saona, L. E. and M. E. Allendorf. 2011. Use of FTIR for rapid authentication and detection of adulteration of food. Annu Rev Food Sci T 2:467-483.

Roggo, Y., P. Chalus, L. Maurer, C. Lema-Martinez, A. Edmond, and N. Jent. 2007. A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies. J Pharmaceut Biomed 44(3):683-700.
Rutten, M. J. M., H. Bovenhuis, J. M. L. Heck, and J. A. M. van Arendonk. 2011a. Predicting bovine milk protein composition based on Fourier transform infrared spectra. Journal Of Dairy Science 94(11):5683-5690.
Rutten, M. J. M., H. Bovenhuis, J. M. L. Heck, and J. A. M. van Arendonk. 2011b. Prediction of beta-lactoglobulin genotypes based on milk Fourier transform infrared spectra. Journal Of Dairy Science 94(8):4183-4188.
Rutten, M. J. M., H. Bovenhuis, K. A. Hettinga, H. J. F. van Valenberg, and J. A. M. van Arendonk. 2009. Predicting bovine milk fat composition using infrared spectroscopy based on milk samples collected in winter and summer. Journal of Dairy Science 92(12):6202-6209.
Rutten, M. J. M., H. Bovenhuis, and J. A. M. van Arendonk. 2010. The effect of the number of observations used for Fourier transform infrared model calibration for bovine milk fat composition on the estimated genetic parameters of the predicted data. Journal of Dairy Science 93(10):4872-4882.
Santos, P. M., E. R. Pereira-Filho, and L. E. Rodriguez-Saona. 2013a. Rapid detection and quantification of milk adulteration using infrared microspectroscopy and chemometrics analysis. Food Chem 138(1):19-24.
Santos, P. M., E. R. Pereira, and L. E. Rodriguez-Saona. 2013b. Application of hand-held and portable infrared spectrometers in bovine milk analysis. J Agr Food Chem 61(6):1205-1211.
Shi, Y. 2014. Identification Study of Milk Quality and Metamorphic Grade Based Spectroscopic Techniques. Chongqing University of Technology, Chongqing.
Sivakesava, S. and J. Irudayaraj. 2002. Rapid Determination of Tetracycline in Milk by FT-MIR and FT-NIR Spectroscopy. Journal of Dairy Science 85(3):487-493.
Soyeurt, H., C. Bastin, F. G. Colinet, V. M. R. Arnould, D. P. Berry, E. Wall, F. Dehareng, H. N. Nguyen, P. Dardenne, J. Schefers, J. Vandenplas, K. Weigel, M. Coffey, L. Theron, J. Detilleux, E. Reding, N. Gengler, and S. McParland. 2012. Mid-infrared prediction of lactoferrin content in bovine milk: potential indicator of mastitis. Animal 6(11):1830-1838.
Soyeurt, H., D. Bruwier, J. M. Romnee, N. Gengler, C. Bertozzi, D. Veselko, and P. Dardenne. 2009. Potential estimation of major mineral contents in cow milk using mid-infrared spectrometry. Journal of Dairy Science 92(6):2444-2454.
Soyeurt, H., F. G. Colinet, V. M. R. Arnould, P. Dardenne, C. Bertozzi, R. Renaville, D. Portetelle, and N. Gengler. 2007. Genetic variability of lactoferrin content estimated by mid-infrared Spectrometry in bovine milk. Journal Of Dairy Science 90(9):4443-4450.
Soyeurt, H., P. Dardenne, F. Dehareng, G. Lognay, D. Veselko, M. Marlier, C. Bertozzi, P. Mayeres, and N. Gengler. 2006. Estimating fatty acid content in cow milk using mid-infrared spectrometry. Journal of Dairy Science 89(9):3690-3695.

Soyeurt, H., F. Dehareng, N. Gengler, S. McParland, E. Wall, D. P. Berry, M. Coffey, and P. Dardenne. 2011. Mid-infrared prediction of bovine milk fatty acids across multiple breeds, production systems, and countries. Journal of Dairy Science 94(4):1657-1667.

Toffanin, V., M. De Marchi, N. Lopez-Villalobos, and M. Cassandro. 2015. Effectiveness of mid-infrared spectroscopy for prediction of the contents of calcium and phosphorus, and titratable acidity of milk and their relationship with milk quality and coagulation properties. Int Dairy J 41:68-73.

Tsenkova, R., S. Atanassova, Y. Ozaki, K. Toyoda, and K. Itoh. 2001. Near-infrared spectroscopy for biomonitoring: influence of somatic cell count on cow's milk composition analysis. Int Dairy J 11(10):779-783.

Tsenkova, R., S. Atanassova, K. Toyoda, Y. Ozaki, K. Itoh, and T. Fearn. 1999. Near-Infrared Spectroscopy for Dairy Management: Measurement of Unhomogenized Milk Composition. Journal of Dairy Science 82(11):2344-2351.

Wang, M., F. Yu, and X. Liu. 2015. Detection of protein, fat in milk by using near infrared diffuse reflectance technique. Laser Journal 36(1):70-73.

Wu, D., Y. He, and S. Feng. 2008. Short-wave near-infrared spectroscopy analysis of major compounds in milk powder and wavelength assignment. Anal Chim Acta 610(2):232-242.

Yang, J., D. Bu, and J. Wang. 2013. Determination of chemical composition of milk using near-infrared reflectance spectroscopy. Food Science 34(20):153-156.

Yang, L. 2011. Study of Infrared Model on Rapid Detection of Five Fatty Acids in Milk Fat. Shandong University of Technology, Zibo.

Yang, R., R. Liu, and K. Xu. 2012. Detection of urea in milk using two-dimensional correlation spectroscopy and partial least square method. Transactions of the Chinese Society of Agricultural Engineering 28(6):259-263.

Yazdanpanah, N. and T. A. G. Langrish. 2013. Comparative study of deteriorative changes in the ageing of milk powder. Journal of Food Engineering 114(1):14-21.

Zhang, H. 2010. Rapid Detection of Milk Protein and Fat Content Using Near Infrared Spectroscopy. Vol. PhD. Central South University, Changsha.

Zhang, L.-G., X. Zhang, L.-J. Ni, Z.-B. Xue, X. Gu, and S.-X. Huang. 2014. Rapid identification of adulterated cow milk by non-linear pattern recognition methods based on near infrared spectroscopy. Food Chem 145:342-348.

Zhao, M., Y. Shi, and B. Luo. 2014. Testing the amount of protein and fat contained in the milk based on near in frared transmission spectroscopy. Laser Journal 35(1):44-50.

Zou, X. B., J. W. Zhao, M. J. W. Povey, M. Holmes, and H. P. Mao. 2010. Variables selection methods in near-infrared spectroscopy. Anal Chim Acta 667(1-2):14-32.

5

Feasibility of detection of milk adulterated with hydrolysed plant protein using mid-infrared spectra or predicted non-protein nitrogen content

Mid-infrared (MIR) spectroscopy reflects fundamental vibrations, deformations, elongations, twistings… of molecular groups, which were used to identify samples adulterated with hydrolyzed plant (soy, rice, and wheat) protein and whey. In order to avoid untargeted fat interference on sample spectral data, skimmed samples were prepared. In addition, estimated non-protein nitrogen (eNPN) of samples, subtracting true protein reading from total protein values, were also calculated to pinpoint adulteration. There was only one mixed commercial milk used as control sample, which means sample spectral variation among authentic milk was not considered in this part. Therefore, the chapter relates a preliminary study to investigate the feasibility of MIR spectroscopy and eNPN measurement to make sure that spectral variations caused by protein difference between adulterated samples could be measured. In this Chapter, detection limit of MIR was the minimum level of qualitative model classified. For the classification model developed using eNPN data, the threshold value was determined.

From Yang, J., N. Zheng, V. Baeten, X. Guo, Y. Yang, J. Wang, and H. Soyeurt. Feasibility of detection of milk adulterated with hydrolysed plant protein using mid-infrared spectra or predicted non-protein nitrogen content. Submitted manuscript.

Abstract:

Hydrolysed plant proteins are considered as candidates for milk adulteration due to their solubility in aqueous phases. The objective of this study is to identify milk adulteration with hydrolysed plant protein via mid-infrared (MIR) spectroscopy or MIR estimated non-protein nitrogen (eNPN). Milk samples adulterated with different levels (30, 15, 7.5, 3.75, 1.88 g/L) of hydrolysed plant protein [soy (HSP), rice (HRP), wheat (HWP)] or whey were prepared for this study. Crude and true protein contents were determined using the Kjeldahl method and a MIR analyser. Spectral data were collected for 21 skimmed milk samples (including adulterated and control samples) using Attenuated Total Reflectance Fourier Transform Infrared Spectroscopy (ATR-FTIR). Estimated NPN (eNPN) was calculated by subtracting true protein from crude protein predicted by MIR. Spectral data point clusters revealed by principal component analysis depended on the level and type of the milk adulterants used. Partial Least Square - Discriminant Analysis (PLS-DA) showed favourable classification of adulterated samples, with specificity of 87.5-100% and sensitivity of 80-100%, only the control milk was misclassified. Subsequent PLS regression models showed good performance of the MIR model on the prediction of adulterant levels (R^2 = 0.95, RMSE = 2.25 g/L of full cross-validation) and eNPN (R^2 = 0.70, RMSE = 0.06 g/100 g of full cross-validation). The addition of adulterants to milk increased the eNPN level for all samples, ranging from 0.04 to 0.37 g/100 g. Combined with the NPN range in normal milk and the predicted eNPN difference of adulterated samples from the control, it was possible to detect samples adulterated with HRP, HSP, or HWP $\geq$ 3.8 g/L. This study indicates the feasibility of detection of milk adulteration with hydrolysed plant protein using MIR or corresponding eNPN prediction, as well as a higher accuracy of classification using MIR.

Key words: Milk adulteration, non-protein nitrogen, Hydrolysed plant protein, Mid-Infrared Spectroscopy

1. Introduction

Milk protein adulteration is the addition of cheap nitrogen-rich compounds to milk to increase the apparent protein content. Milk adulterated with melamine, urea, and whey protein are common cases because these adulterants have high nitrogen or protein contents (Nascimento et al., 2017). Unfortunately, the addition of foreign protein to milk can have detrimental impacts on human health. For instance, the intake of milk powder spiked with melamine causes the occurrence of serious kidney stones, especially for children (Zhang et al., 2014). Additionally, natural milk contained 180-400 mg/L of urea; however, the consumption of urea above 700 mg/L may cause disorders of digestive and excretory systems (Jha et al., 2015). Some milk allergies can be induced by the addition of whey protein, which is rich in α-lactalbumin and β-lactoglobulin that are responsible for these allergies (Chen et al., 2014). Another potential candidate adulterant is hydrolysed plant protein, which includes acidic or enzymatic hydrolysates of high protein plant tissues. These compounds contain a large proportion of free amino acids and peptides, giving them good water solubility (Aaslyng et al., 1998). Although some studies have shown no evidence that hydrolysed plant protein would induce allergic reactions in people who are sensitive to soy or wheat (Reuter et al., 2010), the unlabelled addition of hydrolysed wheat or soy protein to food is not allowed for allergen labelling (Council, 2011).

The Kjeldahl and Dumas methods, which are the current reference methods used to quantify the protein content of milk, unfortunately cannot distinguish foreign nitrogen compounds from authentic milk protein (Nascimento et al., 2017). Therefore, various detection methods have been developed to identify adulteration. For instance, spectroscopy, biosensors, and chromatographic techniques have been investigated (Nascimento et al., 2017; Poonia et al., 2017). The detection of hydrolysed plant proteins seems to be more problematic. Previous studies have found that ultra-high-performance liquid chromatography (UHPLC) systems equipped with UV detection, and sodium dodecyl sulfate (SDS) capillary electrophoresis (CE) have failed to detect hydrolysed wheat protein (at levels below 10%) and hydrolysed soy protein adulterated milk powder, respectively (Lopez-Tapia et al., 1999; Jablonski et al., 2014). However, even if such reference chemical analysis would work, these methods are not suitable for screening purposes due to their cost, the analysis time, and the requirement for qualified staff. For these reasons, an alternative must be found.

One alternative method could be mid-infrared (MIR) spectroscopy which is sensitive to variation of milk protein. There are specific absorbances of milk proteins on the MIR spectrum. For instance, amide I and II bands are located in the 1 500-1 700 cm^{-1} range and the covalent band at 1 060-1 100 cm^{-1} (Etzion et al., 2004). The application of attenuated total reflectance (ATR)-MIR could improve the detection of milk protein adulteration, where a multifold increase in the sample's response is compared with single refraction crystals (Rodriguez-Saona and Allendorf, 2011). When combined with the correct multivariate algorithm, such as polynomial partial least square version (Poly-PLS), artificial neural network (ANN), support vector regression (SVR), and least squares support vector machine (LS-SVM), ATR-MIR was successfully used to detect low concentrations of melamine added to liquid milk

(< 14.6 mg/kg), with a root mean square error for prediction (RMSEP) less than 0.5 mg/kg (Balabin and Smirnov, 2011). From PLS regression, Santos et al. (2013) developed a model to predict the content of whey and urea added to milk, using FT-MIR for 135 samples per adulteration (5 levels × 9 lots × 3 replicates). The obtained standard errors of prediction (SEP) 2.33 g/L for whey and 0.30 g/L for urea, and validation coefficients of determination (R^2v) were 0.96 for whey and 0.98 for urea. More recently, using 210 ATR-MIR spectra (6 levels and 1 control), Jha et al. (2015) developed a model based on soft independence modelling of class analogy (SIMCA) to discriminate milk from milk spiked with urea (100-900 mg/kg and 1 300-2 000 mg/kg). The obtained classification efficiency was higher than 80%. This discrimination was based on the spectral difference in the stretching region between 1 670 and 1 564 cm^{-1}. However, to our knowledge, no studies have investigated the detection of hydrolysed plant protein adulterated milk using ATR-MIR. Therefore, the first objective of this study was to explore the feasibility of using ATR-MIR to qualitatively and quantitatively detect milk adulteration with hydrolysed plant protein.

Another efficient way to monitor milk protein adulteration is to detect the non-protein nitrogen (NPN) levels, estimated by subtracting the casein and whey protein contents (also called true protein, TP) from the crude protein (CP) (Gao et al., 2015; DeVries et al., 2017). The NPN fraction in milk, including urea, ammonia, and free amino acids, ranged from 0.162 to 0.255%, which is much less than CP contents (2.88-4.19%) (Ruska and Jonkus, 2014). Therefore, a level of NPN in milk over this range could indicate milk protein adulteration. The ratio of NPN to crude protein (NPN index) was effectively used by Gao et al. (2015). These authors observed a higher NPN index for raw milk adulterated with melamine (above 0.2 mg/kg) or urea, ammonium chloride, and ammonium sulfate (0.2%, w/w). NPN indices for those adulterated samples were above 1.60, and 5.05-5.42, respectively; while the values for the control milk ranged between 0.81 and 1.50. Currently, FT-MIR spectrometers used for milk testing predict the TP content (Lynch et al., 2006). The estimated quantity of NPN (eNPN) in milk can be found by subtracting milk TP predicted by MIR analysis from CP measured by the Kjeldahl method. Due to the potential of eNPN, the second objective of the current study was to test the feasibility of predicted eNPN to detect the plant protein adulterated milk samples. To our knowledge, this method has not yet been tested by other research teams.

2. Materials and methods

2.1. Sample preparation

Ten retailed milk samples were obtained from Sanyuan Foods (Beijing, China), and a mixture of these milks was prepared as the control milk. Samples were spiked with hydrolysed wheat protein (HWP, CP100, Conpro, Kangke Food Engineering Tech LTD, Wuxi, Jiangsu, China), hydrolysed rice protein (HRP, Shuaixing, Yonggudanbaifen LTD, Wuhan, Hubei, China), and hydrolysed soy protein (HSP, LP0044, Oxoid LTD, Basingstoke, Hampshire, England), as well as whey powder (Xierma, Tongzhougongji Bio-Tech LTD, Minhang district, Shanghai, China).

Protein contents for HWP, HRP, HSP and whey were 71.47 ± 0.12, 56.08 ± 0.26, 49.77 ± 0.09, and 79.93 ± 0.30 g/100g, determined in triplicate using a KjelROC analyser (KD310, OPSIS AB Inc., Sweden). Adulterant-milk solutions of 30 g/L for hydrolysed plant protein and whey protein were homogenised (JJ-2B homogeniser, 200W, 50Hz, 220V, Ronghua Instrument Manufacturing Co. LTD, Changzhou, Jiangsu, China) at 8,000 rpm for 5 min. The subsequent successive gradient dilution process ranged from 6.25 to 50% (v/v), following Santos et al. (2013). So, milk samples were adulterated using the four studied adulterants with 5 different levels (30, 15, 7.5, 3.75, 1.875 g/L). Therefore, 21 samples were prepared in total, including control milk.

2.2. Measurement from whole milk

The CP contents of milk samples were determined in triplicate using a KjelROC analyser (KD310, OPSIS, Furulund, Sweden). TP contents of samples were determined in duplicate using a FT-MIR spectrometer (MilkoScan FT120, FOSS A/S, Hillerød, Denmark). The contents of eNPN were calculated by subtracting average TP from average CP. To evaluate the influence of adulterants on the CP and TP contents of samples, regression lines were drawn between dependent CP and TP values and independent adulterant levels, and corresponding slopes were used to describe the changes of CP and TP in studied samples with increasing adulterant levels in milk.

2.3. Spectral acquisition

In order to avoid scattering caused by milk fat globules, milk samples were skimmed according to the method proposed by Santos et al. (2013). In brief, equal volumes of sample and chloroform were mixed by vortex shaking to remove milk fat. The supernatant after centrifugation was collected for spectra scanning.

Mid-infrared spectra of skimmed samples were acquired by a Fourier Transform Infrared spectrometer (Tensor 27, Bruker optics, MA, US) in attenuated total reflectance (ATR) mode, equipped with a ZnSe cell. Three drops of skimmed samples were placed on the ZnSe crystal. For each individual spectrum an average of 64 scans were carried out at 4 cm^{-1} resolution and over 4 000-650 cm^{-1} range. Between each spectrum acquisition, the crystal cell was wiped dry using soft tissue paper and cleaned by wiping with an alcohol pad. This procedure was repeated 20 times for each adulterated sample and 38 times for the control milk sample to ensure good repeatability. The background spectra of ethanol were recorded at an interval of 20 independent spectra. Finally, all independent spectra from the same adulterated sample were averaged. The final dataset contained 21 spectra.

2.4. Multivariate Data Analysis

Spectra were normalized by subtracting the mean and dividing by the standard deviation of each individual spectra. These normalized spectra were then exported to R software (version 3.4.1, Bell Laboratories, New Jersey, USA). The library caret (version 6.0-78) was used to realise the multivariate analysis. Principle component analysis (PCA) was used to derive the sample clustering according to adulterant type

and level. The score and loading plot of the first two principal components were exported to show the sample distribution. Partial Least Square - discriminate analysis (PLS-DA) with a leave-one-out cross-validation was carried out to identify the type of adulterant in the studied samples. Metrics of classification model, such as sensitivity, specificity, precision, and accuracy, were exported to explain the model performances in adulterant identification, as well as the confusion matrix to assess the discernment ability of the developed models. The VIP score of PLS-DA and the decision tree scheme were used to clarify the important variables of the corresponding model. A PLS regression with a leave-one-out cross-validation was also performed to predict the quantity of adulterant in the studied milk. Moreover, a PLS regression model to predict eNPN contents of samples was developed. Standard error (SEC) and coefficients of determination of calibration (R^2c) and of cross-validation (R^2cv) were calculated to assess the robustness of the developed regression models.

Table 5-1: Average and standard deviation of crude protein (CP) measured by Kjedahl (g/100 g of milk), and true protein (TP) determined by MIR spectrometer (g/100 g of milk) following the studied adulterant and level (g/L of milk).

	Adulterant level	Protein contents (g/100 g of milk)			
	(g/L of milk)	HRP	HSP	HWP	Whey
CP	30.0	4.78 ± 0.01	4.70 ± 0.00	5.42 ± 0.01	5.45 ± 0.03
	15.0	4.03 ± 0.03	3.95 ± 0.00	4.29 ± 0.01	4.32 ± 0.04
	7.5	3.70 ± 0.00	3.59 ± 0.00	3.75 ± 0.00	3.80 ± 0.03
	3.8	3.50 ± 0.01	3.41 ± 0.01	3.49 ± 0.02	3.52 ± 0.00
	1.9	3.38 ± 0.01	3.31 ± 0.01	3.32 ± 0.01	3.40 ± 0.05
	Slope	*0.05*	*0.05*	*0.07*	*0.07*
TP	30.0	4.68 ± 0.00	4.39 ± 0.00	5.04 ± 0.01	5.42 ± 0.01
	15.0	3.96 ± 0.01	3.78 ± 0.00	4.11 ± 0.01	4.33 ± 0.00
	7.5	3.60 ± 0.00	3.51 ± 0.00	3.64 ± 0.01	3.78 ± 0.00
	3.8	3.41 ± 0.00	3.36 ± 0.01	3.44 ± 0.00	3.50 ± 0.00
	1.9	3.31 ± 0.01	3.30 ± 0.01	3.33 ± 0.00	3.35 ± 0.00
	Slope	*0.05*	*0.04*	*0.06*	*0.07*

Notes: HRP = hydrolysed rice protein; HSP = hydrolysed soy protein; HWP = hydrolysed wheat protein. Slopes indicate the changed protein values (g/100 g) of samples with the per unit (g/L) addition of adulterants.

3. Results and Discussion

3.1. Sample protein contents

The sample values for CP measured by Kjeldahl and TP predicted by MIR spectrometry are listed in Table 5-1. The CP and TP contents in control milk are 3.17 and 3.22 g/100 g, respectively. Independent of the adulterant type, the inclusion of hydrolysed plant protein or whey in milk increased the content of TP (3.30-5.42 g/100 g) and CP (3.31-5.45 g/100 g). However, the increments of TP and CP were different. The addition of whey and HWP increased CP by 0.07 units (1 unit

equals 1 g/100 g of milk protein per 1 g/L of adulterant level), while addition of HRP and HSP raised CP by 0.05 units. The TP increments caused by the addition of HSP, HRP, HWP, and whey were 0.04, 0.05, 0.06 and 0.07 units, respectively.

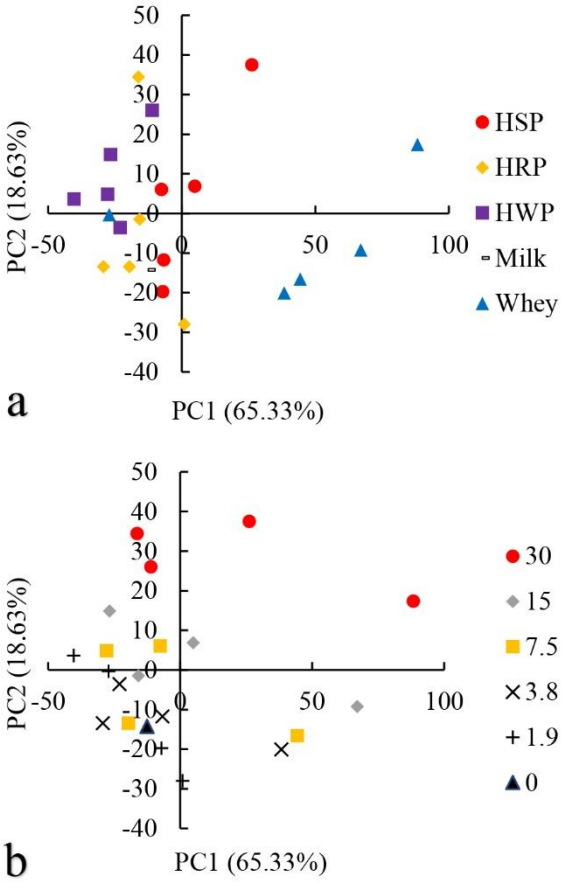


Figure 5-1: Score plots of spectra data by adulterants added to samples (A) and levels (g/L) of adulterants (B) derived from principle component analysis of different levels of milk adulteration. Notes: HRP = hydrolysed rice protein; HSP = hydrolysed soy protein; HWP = hydrolysed wheat protein.

3.2. Identification and quantification of milk adulteration using MIR

Score plots of normalised spectral data from PCA are presented in Figure 5-1. The first two principal components (PCs) accounted for 83.96% of the total variance. Score plots of the first 2 PCs from PCA showed potential sample clusters by adulterant type and level. Based on the score plots, PC1 allowed a partial discrimination of samples according to adulterant (Figure 5-1a). Milk samples spiked with whey were clearly separated by PC1 as they showed the highest values. Then, HSP milk was

separated and was located between adulterated whey samples and the remaining adulterated samples. However, the distinction between control milk, HRP and HWP milk was not possible based on PC1 scores. Adulterated samples with high levels of adulterant (30 g/L) gathered at the positive values of the PC2 axis (Figure. 5-1b). There was a gradual transition of samples located in the downward direction of PC2 as a function of the adulteration level, but no visible boundaries were observed between samples with adulteration levels below 30 g/L.

Table 5-2: Confusion matrix of PLS-DA with leave-one-out cross validation

		Reference value				
		Milk	HRP	HSP	HWP	Whey
Predicted value	Milk	0	0	0	0	0
	HRP	0	4	0	0	0
	HSP	1	1	5	0	0
	HWP	0	0	0	5	1
	Whey	0	0	0	0	4

Notes: HRP = hydrolysed rice protein; HSP = hydrolysed soy protein; HWP = hydrolysed wheat protein.

Table 5-3: Performance of the PLS-DA model on each type of milk adulteration with leave-one-out cross validation

Adulterants	HRP	HSP	HWP	Whey
Sensitivity	80.0%	100%	100%	80.0%
Specificity	100%	87.5%	93.8%	100%
Precision	100%	71.4%	83.3%	100%
Accuracy	90.0%	93.8%	96.9%	90.0%

Notes: HRP = hydrolysed rice protein; HSP = hydrolysed soy protein; HWP = hydrolysed wheat protein.

In order to confirm the first PCA observations of the feasibility of MIR to detect milk samples adulterated with hydrolysed plant protein or whey, a PLS-DA analysis was performed. Based on cross-validation results, 3 factors were included in the PLS-DA model. The cross-validation confusion matrix and corresponding statistical parameters are given in Table 5-2 and 5-3, respectively. These results suggest that the PLS-DA model was more sensitive to adulteration with HSP and HWP. The specificity and accuracy for all of the studied adulterants were higher than 85% and 90%, respectively. These parameters are comparable to the models' performance with other adulterants, as published in recent literature (Botelho et al., 2015; Coitinho et al., 2017; Gondim et al., 2017), in which both sensitivity and specificity ranged from 80% to 100%. Based on the results of the confusion matrix (Table 5-2), nearly all samples were well classified, except 3 samples: the control milk, one sample with HRP, and one sample with whey were classified as HSP, HSP and HWP samples,

respectively. The misclassified HWP and whey samples had a low level of adulterant (1.9 g/L). The VIP scores exported from the developed PLS-DA model allowed for interpretation of the regions responsible for the obtained classification. As plotted in Figure 5-2, high VIP scores (> 40) occurred at 600-640, 951-976, 1 468-1 518, 1 688-1 996, 2 361-2 369, and 3 746-3 757 cm^{-1}. High scores at 600-640 cm^{-1} are associated with out-of-plane bending of N-H and C=O, while high weight at 1 468-1 518 cm^{-1} and 1 688-1 996 cm^{-1} are related to absorption of amide II and I (Carbonaro and Nucara, 2010). The VIP score peaks centred at 2 365 and 3 750 cm^{-1} are associated with the absorption of CO_2 and water, respectively (Aernouts et al., 2011).

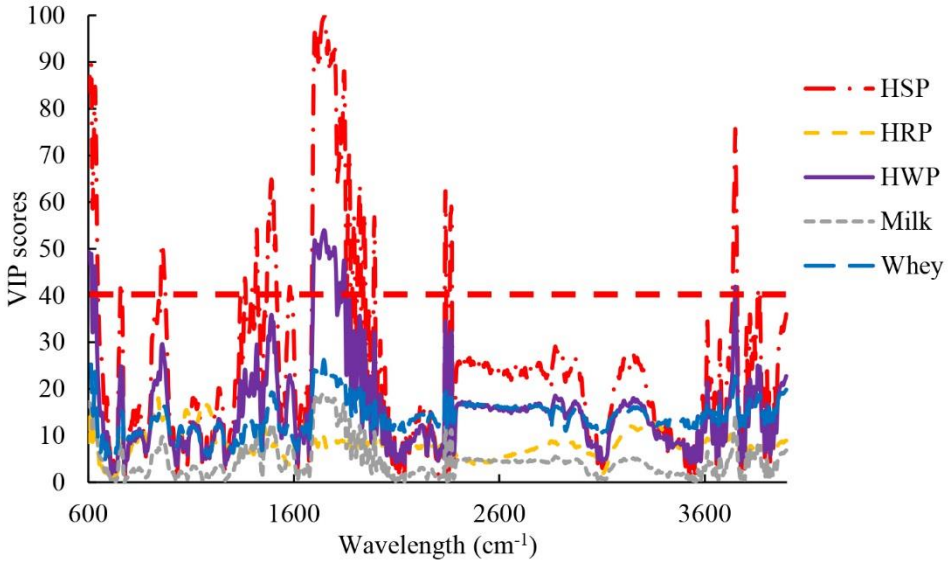


Figure 5-2: Variable importance in project (VIP) scores for the PLS-DA models detecting adulteration. Notes: HRP = hydrolysed rice protein; HSP = hydrolysed soy protein; HWP = hydrolysed wheat protein.

Another objective of this study was also to quantitatively detect the presence of adulterant in milk, independent of the adulterant type. Therefore, PLS regression based on adulterant level was also performed. The obtained PLS regression plot (Figure 5-3a) showed a good fit between the prediction and reference, with standard error = 0.65 g/100 g, R^2 = 0.998 of calibration, and R^2 = 0.952 of cross validation.

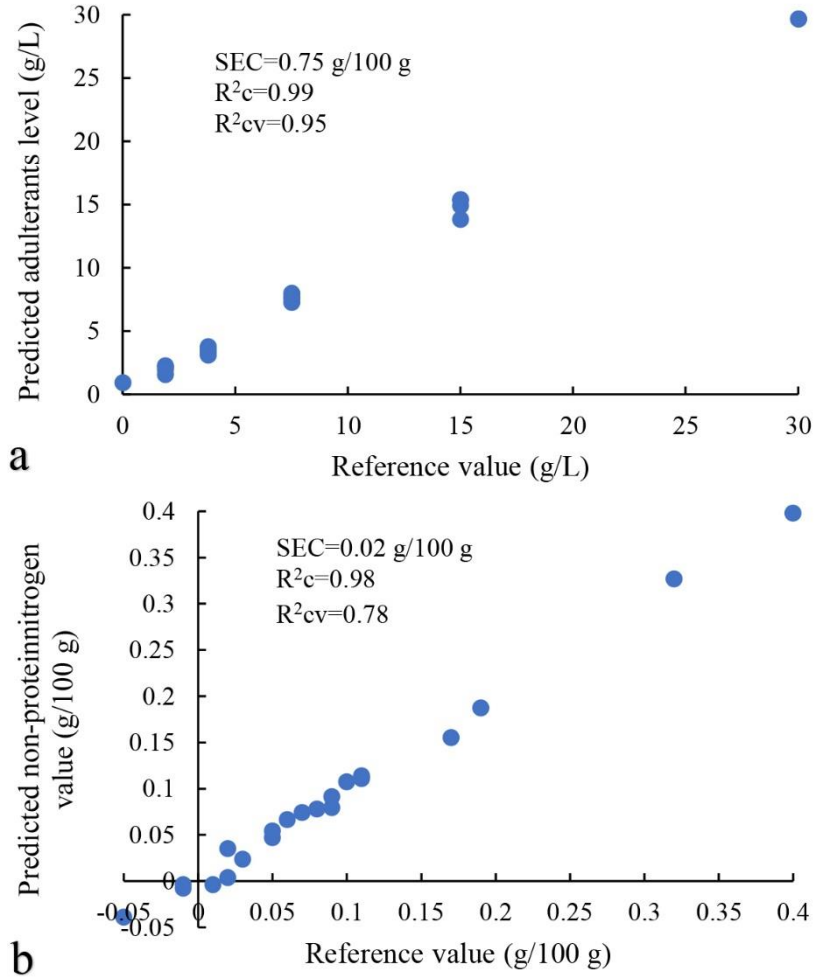


Figure 5-3: Plots of reference vs. predicted values and metrics for calibration and cross-validation of each partial least square (PLS) regression model: a, adulteration level; b, estimated non-protein nitrogen (eNPN).

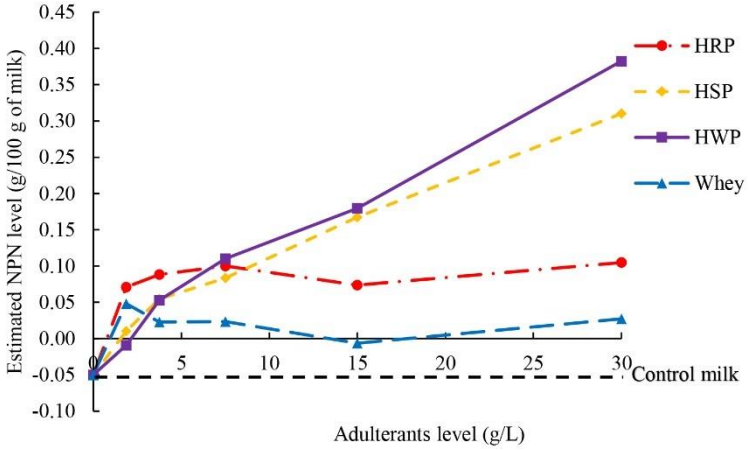


Figure 5-4: Estimated non-protein nitrogen (eNPN) levels in different samples of adulterated milk. HRP = hydrolysed rice protein; HSP = hydrolysed soy protein; HWP = hydrolysed wheat protein.

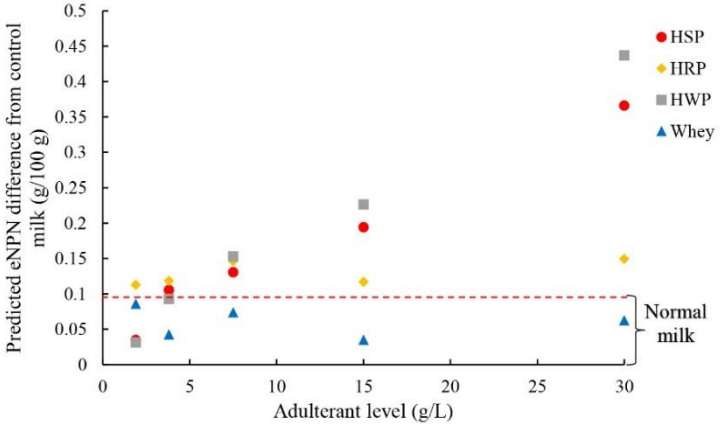


Figure 5-5: Predicted estimated non-protein nitrogen (eNPN) difference for each sample from control milk as a function of the levels and types of adulterants in milk adulteration. The NPN threshold for normal milk is 0.093 g/100 g, suggested by Ruska and Jonkus (2014). HRP = hydrolysed rice protein; HSP = hydrolysed soy protein; HWP = hydrolysed wheat protein.

3.3. Identification of milk adulteration using eNPN predictions

The eNPN contents in samples are presented in Figure 5-4. Nearly all adulterated samples have a higher level (–0.01-0.38 g/100 g) than control milk (–0.05 g/100 g). An increase in eNPN in samples was observed when HSP and HWP adulteration

increased. This is related to the high content of dissolved NPN (Aaslyng et al., 1998) from the hydrolysed plant proteins used. However, a more or less flat tendency with an average content of 0.09 g of eNPN/100 g of milk was observed for HRP samples. The constant eNPN contents for HRP adulteration may indicate a low released level of NPN (0.07-0.10 g/100 g) in HRP adulterated milk samples. A similar flat tendency was also observed for whey adulteration, with a lower content (0.02 g of eNPN/100 g of milk). This can be explained by the low content of NPN in whey protein powder (less than 1%) (Svanborg et al., 2015). Three negative values of eNPN content were obtained: one in control milk (–0.05 g/100 g) and one for samples adulterated with whey and HWP at 1.9 g/L (–0.01 g/100 g). The eNPN level of the control sample in this study is lower than the NPN range (0.162-0.255 g/100 g) reported in other studies (Ruska and Jonkus, 2014). Corresponding biases may result from overestimation of TP by the FT-MIR spectrometer.

Although the detection of NPN has been used for the identification of milk adulteration (Gao et al., 2015), Kjeldahl measurements of CP and TP in milk take time. Therefore, it could be faster to directly predict eNPN using MIR. The performance of eNPN prediction using PLS regression is shown in Figure 5-3b. Values of R^2 and RMSE for cross validation of eNPN were equal to 0.70 and 0.06 g/100 g, respectively.

Combined with the NPN range of normal milks, the predicted eNPN would be useful for the identification of milk adulteration. As reported by Ruska and Jonkus (2014), NPN varied from 0.162 to 0.255 g/100 g in normal milk. In other words, when the predicted eNPN gap between sample and control milk exceeds the NPN range in control milk, 0.093 g/100 g, the sample would be classified as adulterated. Alternatively, calculation of the difference of the sample eNPN from control milk could reduce potential bias caused by overestimation of TP in this study.

Figure 5-5 presents predicted eNPN difference plots for each sample compared to control milk, only samples with predicted eNPN difference (sample minus control milk) less than 0.093 g/100 g are recognised as normal. Therefore, whey samples, HSP and HWP samples at the level of 1.9 g/L are identified as normal milk. It should be noted that all samples adulterated with whey could not be recognised by the developed model. As discussed above, amounts of NPN released from whey powder in adulterated samples is limited (0.05 g/100 g). The NPN proportion for adulteration with whey would be very low, even less than the control milk (Gao et al., 2015). Therefore, this model may not be sensitive to milk adulteration with whey. Otherwise, NPN contents in milk products are sensitive to adulteration with ammonium, urea, melamine, and other nitrogen-rich substances (Gao et al., 2015; DeVries et al., 2017).

Compared with the developed classification model using spectra data, the predicted eNPN model is not so efficient for the recognition of milk adulteration with whey and low level (1.9 g/L) HSP and HWP, which means that for the detection of milk adulteration a classification model directly using spectral data is more suitable than calculated parameters using spectra predictions in this study. This is in line with efficiency comparisons of direct and indirect models for the estimated breeding value of milk composition (Dagnachew et al., 2013).

The eNPN model in the current study needs to be improved and validated in future

studies: (1) The spectral variance of normal milk should be taken into consideration; more NPN values for normal milk samples should be added to the database, which primarily affects the threshold value of the developed eNPN model. (2) Spectral and NPN data for adulteration with urea, melamine, and other nitrogen-rich substances in milk should be included. (3) As a potential identification model for non-targeted adulteration, the composition of the NPN fraction recognised by this model, such as pure chemicals, free amino acids, or small peptides, should be defined.

4. Conclusion

In this study, we evaluated the potential for the detection of milk adulteration with hydrolysed plant protein in 2 ways: a MIRS model of identification and quantification of milk adulteration, and predicted eNPN to identify adulteration. The developed classification model using a decision tree could correctly classify all the adulterated samples according to type of adulterant, and successive PLS regression models could quantify the adulterant level with high accuracy (R^2cv = 0.95, RMSEcv = 2.25 g/L). Combined with reported NPN range (0.162-0.255 g/100 g) in unadulterated milk and predicted eNPN differences from control milk in samples using a MIRS model, samples with HRP at all levels, and HSP and HWP at levels above 3.8 g/L are classified as adulterated samples.

5. Acknowledgements

This study was supported by the Special Fund for Agro-scientific Research in the Public Interest (201403071), the Agricultural Science and Technology Innovation Program (ASTIP-IAS12), Modern Agro-Industry Technology Research System of the PR China (CARS-36), and the Project of Risk Assessment on Raw Milk (GJFP2018008).

6. References

Aaslyng, M. D., M. Martens, L. Poll, P. M. Nielsen, H. Flyge, and L. M. Larsen. 1998. Chemical and sensory characterization of hydrolyzed vegetable protein, a savory flavoring. J Agric Food Chem 46(2):481-489.

Aernouts, B., E. Polshin, W. Saeys, and J. Lammertyn. 2011. Mid-infrared spectrometry of milk for dairy metabolomics: a comparison of two sampling techniques and effect of homogenization. Anal Chim Acta 705(1-2):88-97.

Balabin, R. M. and S. V. Smirnov. 2011. Melamine detection by mid- and near-infrared (MIR/NIR) spectroscopy: a quick and sensitive method for dairy products analysis including liquid milk, infant formula, and milk powder. Talanta 85(1):562-568.

Botelho, B. G., N. Reis, L. S. Oliveira, and M. M. Sena. 2015. Development and analytical validation of a screening method for simultaneous detection of five adulterants in raw milk using mid-infrared spectroscopy and PLS-DA. Food Chem 181:31-37.

Carbonaro, M. and A. Nucara. 2010. Secondary structure of food proteins by Fourier transform spectroscopy in the mid-infrared region. Amino Acids 38(3):679-690.
Chen, F.-M., J.-H. Lee, Y.-H. Yang, Y.-T. Lin, L.-C. Wang, H.-H. Yu, and B.-L. Chiang. 2014. Analysis of α-lactalbumin-, β-lactoglobulin-, and casein-specific IgE among children with atopic diseases in a tertiary medical center in northern Taiwan. Journal of Microbiology, Immunology and Infection 47(2):130-136.
Coitinho, T. B., L. D. Cassoli, P. H. R. Cerqueira, H. K. da Silva, J. B. Coitinho, and P. F. Machado. 2017. Adulteration identification in raw milk using Fourier transform infrared spectroscopy. J Food Sci Technol 54(8):2394-2402.
Council, E. P. a. 2011. Regulation (EU) No 1169/2011 of the European Parliament and of the Council of 25 October 2011 on the provision of food information to consumers, amending Regulations (EC) No 1924/2006 and (EC) No 1925/2006 of the European Parliament and of the Council, and repealing Commission Directive 87/250/EEC, Council Directive 90/496/EEC, Commission Directive 1999/10/EC, Directive 2000/13/EC of the European Parliament and of the Council, Commission Directives 2002/67/EC and 2008/5/EC and Commission Regulation (EC) No 608/2004. Pages 18-63 in Off. J. Eur. Union. Vol. 50.
Dagnachew, B. S., T. H. E. Meuwissen, and T. Adnoy. 2013. Genetic components of milk Fourier-transform infrared spectra used to predict breeding values for milk composition and quality traits in dairy goats. Journal Of Dairy Science 96(9):5933-5942.
DeVries, J. W., G. W. Greene, A. Payne, S. Zbylut, P. F. Scholl, P. Wehling, J. M. Evers, and J. C. Moore. 2017. Non-protein nitrogen determination: A screening tool for nitrogenous compound adulteration of milk powder. Int Dairy J 68:46-51.
Etzion, Y., R. Linker, U. Cogan, and I. Shmulevich. 2004. Determination of protein concentration in raw milk by mid-infrared fourier transform infrared/attenuated total reflectance spectroscopy. J Dairy Sci 87(9):2779-2788.
Gao, P., Z. Li, L. Zan, T. Yue, and B. Shi. 2015. A non-protein nitrogen index for discriminating raw milk protein adulteration via the Kjeldahl method. Analytical Methods 7(21):9166-9170.
Gondim, C. d. S., R. G. Junqueira, S. V. C. d. Souza, I. Ruisánchez, and M. P. Callao. 2017. Detection of several common adulterants in raw milk by MID-infrared spectroscopy and one-class and multi-class multivariate strategies. Food Chem 230:68-75.
Jablonski, J. E., J. C. Moore, and J. M. Harnly. 2014. Nontargeted detection of adulteration of skim milk powder with foreign proteins using UHPLC-UV. J Agric Food Chem 62(22):5198-5206.
Jha, S. N., P. Jaiswal, A. Borah, A. K. Gautam, and N. Srivastava. 2015. Detection and quantification of urea in milk using attenuated total reflectance-fourier transform infrared spectroscopy. Food and Bioprocess Technology 8(4):926-933.
Lopez-Tapia, J., M. R. Garcia-Risco, M. A. Manso, and R. Lopez-Fandino. 1999. Detection of the presence of soya protein in milk powder by sodium dodecyl sulfate capillary electrophoresis. J Chromatogr A 836(1):153-160.
Lynch, J. M., D. M. Barbano, M. Schweisthal, and J. R. Fleming. 2006. Precalibration evaluation procedures for mid-infrared milk analyzers. J Dairy Sci 89(7):2761-2774.

Nascimento, C. F., P. M. Santos, E. R. Pereira-Filho, and F. R. P. Rocha. 2017. Recent advances on determination of milk adulterants. Food Chem 221:1232-1244.

Poonia, A., A. Jha, R. Sharma, H. B. Singh, A. K. Rai, and N. Sharma. 2017. Detection of adulteration in milk: A review. International Journal of Dairy Technology 70(1):23-42.

Reuter, A., R. Gerald, Y. Kuehne, A. Engin, J. Przywara, M. Worm, B. Ballmer-Weber, T. Holzhauser, and S. Vieths. 2010. Evaluation of the allergenic potential of soy and wheat based seasonings. Allergy: European Journal of Allergy and Clinical Immunology 65:322.

Rodriguez-Saona, L. E. and M. E. Allendorf. 2011. Use of FTIR for rapid authentication and detection of adulteration of food. Annu Rev Food Sci T 2:467-483.

Ruska, D. and D. Jonkus. 2014. Crude protein and non-protein nitrogen content in dairy cow milk. Proceedings of the Latvia University of Agriculture 32(1):36-40.

Santos, P. M., E. R. Pereira, and L. E. Rodriguez-Saona. 2013. Application of hand-held and portable infrared spectrometers in bovine milk analysis. J Agr Food Chem 61(6):1205-1211.

Svanborg, S., A. G. Johansen, R. K. Abrahamsen, and S. B. Skeie. 2015. The composition and functional properties of whey protein concentrates produced from buttermilk are comparable with those of whey protein concentrates produced from skimmed milk. J Dairy Sci 98(9):5829-5840.

Zhang, L. G., X. Zhang, L. J. Ni, Z. B. Xue, X. Gu, and S. X. Huang. 2014. Rapid identification of adulterated cow milk by non-linear pattern recognition methods based on near infrared spectroscopy. Food Chem 145:342-348.

6

Identification of non-milk protein adulterated in milk using mid-infrared spectroscopy- a case of detection of hydrolyzed plant protein

In this chapter, nine raw milks were designed as control samples, in order to take into account spectral variation between authentic milks. The results were interpreted on the basis of "corrected" spectra excluding absorbances linked to the fat moiety of milk. Hydrolyzed wheat and rice protein, whey, and urea were spiked in milk as foreign protein or nitrogen adulterants. Near- infrared spectroscopy, response of overtone and combination vibrations of chemical bonds, was used to identify the milk adulteration, together with mid-infrared spectroscopy. In model validation, an innovative double validation strategy was adopted in this part. Ten-folds cross-validation was performed to obtain the best performance for each classification and regression model and random whole dataset splitting was replicated to evaluate the variability of statistics parameters. Detection limits of spectroscopy methods in this part was defined as the minimum level of classification model could identify.

From Yang, J., N. Zheng, J. Wang, V. Baeten, and H. Soyeurt. Identification of non-milk protein adulterated in milk using mid-infrared spectroscopy- a case of detection of hydrolyzed plant protein. Submitted paper.

Abstract

This study aims to develop a rapid detection method for milk adulterated with foreign proteins using mid- and near- infrared spectroscopy. A total of 9 raw milk samples were adulterated with 5 levels of hydrolysed rice protein (HRP, 2.5-40 g/L), hydrolysed wheat protein (HWP, 1.88-30 g/L), whey (1.88-30 g/L), urea (0.5-8 g/L), and water (3.13-50 g/L), as a result of a series dilution process (50.0-3.13%). These samples were analysed using attenuated total reflectance mid-infrared (ATR-MIR) and near-infrared (NIR) spectrometers. Raw ATR-MIR and first derivative NIR spectra data showed better classification and regression model performance than other pre-treated data studied. The developed partial least squares-discriminant analysis (PLS-DA) model could not discriminate control milk from adulterated samples, probably due to the compositional variability of control milks. ATR-MIR had better discriminability for HRP and HWP above 6.25%, while NIR showed its better performing discriminability for whey above 12.5%. Similar good regression model performance was seen for quantification of the level of HRP and urea in milk using both ATR-MIR and NIR milk spectra (validation $R^2p > 0.96$). ATR-MIR predicted the HWP level better than NIR (validation $R^2p = 0.95$ vs. 0.88), while NIR had a better accuracy for whey level quantification (validation $R^2p = 0.97$ vs. 0.40). Moreover, both spectroscopies provided a good prediction of the protein adulterants level (excluding water) in adulterated samples (validation $R^2p = 0.87$-0.97, RMSEP = 1.98-4.10 g/L). Consequently, this study shows the complementarity of information provided by NIR (water and whey detection) and MIR (HRP, HWP and urea). However, strong variability was observed between the validation sets used, suggesting a need to increase the size of the calibration sets used to build the models.

Key words: mid-infrared, near-infrared, hydrolyzed plant protein, milk adulteration

1. Introduction

Milk protein adulteration is the addition of foreign nitrogen components to milk or dairy products in order to increase the apparent crude protein content. Melamine, urea, whey, and different plant proteins are often used as candidate adulterants due to their low cost and high nitrogen content (Poonia et al., 2017). The illegal or undeclared addition of such adulterants has negative health consequences. Besides the serious damage to kidneys caused by overdose intakes of melamine and urea (Handford et al., 2016), possible hypotension induced by whey (Renny et al., 2005) and anaphylaxis associated with plant proteins (Ho et al., 2014) could occur. Unfortunately, the chemical reference methods used to quantify the protein content in milk, such as Kjeldahl and Dumas, are not able to distinguish between non-milk protein and true milk protein (Nascimento et al., 2017). Therefore, reliable and accurate methods should be developed to detect these milk adulterations that are potentially dangerous to human health.

As a rapid and non-invasive detection methodology, infrared spectroscopy is based on the vibrations of molecules at different frequencies (Luykx and van Ruth, 2008). Near-infrared (NIR) spectra are the result of overtones and combinations of fundamental vibrations related to chemical groups, such as C-H, N-H, and O-H, generally at high vibrational frequency (4 000-12 500 cm^{-1}) (Rodriguez-Saona and Allendorf, 2011). Mid-infrared (MIR) spectroscopy at 400-4 000 cm^{-1} focuses on the electromagnetic spectrum, monitoring the fundamental vibrational and rotational stretch of molecules in organic compounds (Lohumi et al., 2015). Generally, compared with models developed from dispersive NIR spectra, MIR models using the region located between 700 and 1 200 cm^{-1} have superior performance for quantitative prediction and qualitative classification (Rodriguez-Saona and Allendorf, 2011). Some investigations have already used infrared spectroscopy to detect adulterated milk samples. Santos et al. (2013b) developed a 3D plot of soft independent modelling of class analogies (SIMCA) obtained from a hand-held NIR system (10 000-4 000 cm^{-1}). This plot only separated samples with the highest level ($\geqslant$ 50%) of adulterants (whey, water, synthetic urine, urea, synthetic milk and hydrogen peroxide) from other samples according to a dilution process. The classification performance of validation was 80% for control milk and 56% for adulterated milk. These authors also tested detection using MIR spectroscopy. A portable MIR instrument (1 300-950 cm^{-1}) and a hand-held one (1 800-800 cm^{-1}) provided spectra that allowed good separation between control and adulterated fluid milk samples. The correct classification performance was 100% and 70%, respectively. For the fluid milk adulterated with melamine at a range of 0-14.6 mg/kg and 14.6-2 000 mg/kg, Balabin and Smirnov (2011) obtained similar or superior root mean squared errors (RMSE) of prediction (0.28 and 6.1 mg/kg) from NIR (450-3 850 cm^{-1}) compared with MIR spectroscopy (4 500-9 000 cm^{-1}). According to Jaiswal et al. (2015), the MIR spectral region between 1 680-1 058 cm^{-1} allows the differentiation of milk, soy milk, and milk adulterated with soy milk. This spectral differentiation is related to amides (I, II, III), beta-sheet proteins, α-tocopherol and Kunitz soybean trypsin inhibitor. Based on the spectral differences, the SIMCA model allows classification of control milk versus

adulterated samples at a level of 2-40% with a classification performance above 93% (Jaiswal et al., 2015). Within the spectral range at 1 089-1 058 cm^{-1} and using PLS regression, the soy milk levels in adulterated samples can be predicted with a R^2 of 0.97 and prediction RMSE of 4.59% (Jaiswal et al., 2015). Combined with multiple linear regression, the first derived NIR spectra predict vegetable protein levels (0-5%) in adulterated milk powder, with a standard error (SE) of prediction of 0.23% and a validation SE of 0.21% (Maraboli et al., 2002).

This study aims to investigate the possibility of using MIR and NIR spectroscopy to detect raw milk adulteration, especially with hydrolysed plant protein. Besides the use of hydrolysed plant protein, the innovative aspect of this paper is based on the collection of individual cow milk samples in order to introduce the natural variation in protein contents in control milk samples.

2. Materials and Method

2.1. Milk samples

Table 6-1: Level of adulterant in the adulterated milk samples following the used dilution.

Dilution (% v/v)	50.00	25.00	12.50	6.25	3.13
HRP (g/L of milk)	40.00	20.00	10.00	5.00	2.50
HWP (g/L of milk)	30.00	15.00	7.50	3.75	1.88
Urea (g/L of milk)	8.00	4.00	2.00	1.00	0.50
Whey (g/L of milk)	30.00	15.00	7.50	3.75	1.88
Water (g/L of milk)	50.00	25.00	12.50	6.25	3.13

Notes: HRP = hydrolysed rice protein; HWP = hydrolysed wheat protein

Nine individual milk samples were collected from 9 different cows belonging to 2 local Belgian dairy farms (Gembloux and Jodoigne, Belgium) during February and April 2018. Samples were adulterated with hydrolysed wheat protein (HWP, CP100, Conpro, Kangke Food Engineering Tech LTD, Wuxi, Jiangsu, China), hydrolysed rice protein (HRP, Shuaixing, Yongguodanbaifen LTD, Wuhan, Hubei, China), whey powder (Xierma, Tongzhougongji Bio-Tech LTD, Minhang district, Shanghai, China), urea (Sigma-Aldrich Co., 3050 Spruce Street, St. Louis, MO 63103, USA), and distilled water. Protein contents of HWP, HRP, and whey were 71.47 ± 0.12, 56.08 ± 0.26, and 79.93 ± 0.30 g/100 g, as determined by KjelROC analyzer (KD310, Sweden, OPSIS AB Inc.) from triplicates. The levels of spiked whey and urea used this study (30 and 8 g/L, respectively) were the same as those mentioned in previous reports (Santos et al., 2013a, Santos et al., 2013b). In order to have similar protein levels to whey samples, milk samples with 40 g/L of HRP and 30 g/L of HWP were prepared for this study. Control milk samples were also adulterated by adding water to check spectral changes during the dilution process. The target concentrations of adulterant in milk (40 g/L for HRP, 30 g/L for HWP and whey, 8 g/L for urea, and 50 g/L for water) were prepared by magnetic stirring for 20 min. Calculated crude protein increases in samples induced by the addition of HRP, HWP, urea, and whey were 2.24,

2.14, 2.40, and 2.34 g/100 g respectively. Then, a serial gradient dilution with milk of adulterated samples was followed, as described by Santos et al. (2013b) and ranged from 3.13% to 50.00% v/v. Therefore, five types of adulterated samples (i.e., HWP, HRP, urea, whey and water) at five different levels were created. Table 6-1 shows the links between adulterant level and dilution. The control milk was also included in the set. Therefore, 26 samples were created per analysed milk sample. All samples were then stored at 4°C and warmed to 40°C for 20 min before spectral analysis. The spectroscopic measurements were completed within 24 hours of sample preparation. This sample preparation was repeated 9 times as 9 different milk samples were collected. Consequently, there were a total of 234 samples for spectra scanning.

2.2. Infrared spectroscopy measurements

Mid-infrared spectral measurements were carried out on a Fourier Transform (FT)-MIR spectrometer (Tensor 27, Bruker optics, MA, US) with an attenuated total reflectance (ATR) accessory (GS10500-Z, Specac Ltd, Orpington, Kent, England). Three drops of each well shaken sample were placed on the diamond crystal. Spectra were collected in the region located between 4 000 and 600 cm^{-1} co-adding 64 scans at room temperature (24-25°C). The resolution was 4 cm^{-1}. Four individual spectra (i.e., 4 replicates) were collected for each sample via OPUS software (v. 6.5, Bruker Optics, Ettlingen, Germany), and distilled water was used as the background during sample measurement intervals. After scanning each sample, the ATR crystal was cleaned using soft tissue paper with alcohol and air-dried. All spectra were averaged for further analysis. Therefore, the final data set contained 234 ATR-MIR spectra.

NIR measurements were taken using a FT-NIR spectrometer (MPA, Bruker Optics, Ettlingen, Germany), attached to a liquid sampling module which could pump and homogenise raw milk samples. The sampling apparatus used 40 mL of each sample. Two individual transmitted spectra ranging from 12 500 to 4 000 cm^{-1} were collected with a resolution of 16 cm^{-1}. NIR spectra collection was performed by OPUS software (V6.5, Bruker Optics, Ettlingen, Germany). At the end, NIR spectra were averaged for the further chemometric analysis (N = 234). In order to observe the changes in milk composition, the contents of fat, protein, lactose, and total solids of samples were also predicted using the same spectrometer.

2.3. Chemometric analysis

All data treatments were performed using R software (version 3.4.1, Bell Laboratories, New Jersey, USA). First milk composition changes were assessed using P-values of paired t-tests (package "stats) to compare adulterated vs. control milk using the contents of fat, protein, lactose and total solids predicted by NIR.

Second, principal component analysis of centered and scaled spectra was performed (R package "FactoMineR", version 1.41; Le et al., 2008) to screen potential spectral outliers and to investigate potential samples clusters by adulterant type, level, sampling date, and dilution process. According to Whitfield et al. (1987), a modified Mahalanobis distance (GH) higher than 3 suggests a potential outlier. This distance was calculated for the obtained principal components and only 2 ATR-MIR spectral

data showed a GH higher than 3. These were discarded from the database (N = 232). For the NIR spectral data, 12 NIR spectra were considered to be outliers (N = 220).

Using R package "prospectr" (Version 0.1.3; Stevens and Ramirez-Lopez, 2013), different data processing methods were tested: standard normal variate (SNV), first derivative with a gap of 5 (1D), second derivative with a gap of 5 (2D), as well as combinations of 1D+SNV and 2D+SNV. SNV treatment allowed multiplicative scatter correction in reflectance spectroscopy (Rodriguez-Saona et al., 2000), while first and second derivatives removed additive baseline effects to enhance signal/noise ratio. The combination of these two pretreatments has often been applied in previous studies (Wang et al., 2006, Balabin and Smirnov, 2011).

Classifications based on adulterant type were realised using PLS-discriminant analysis (R package "caret", version 6.0-80; Kuhn et al., 2018). The number of components was fixed when the largest 10-fold cross-validation accuracy was reached (i.e., the ratio of the number of correct classifications to the total number of records). In order to assess the robustness of the obtained classification, the Cohen's kappa unweighted parameter was also calculated. This compared the calculated accuracy with the expected accuracy representing the random chance of having a good classification. Sensitivity and specificity for each adulterant type were also calculated.

Quantification of the level of the adulterants in milk was performed using PLS regression (R package "caret", version 6.0-80; Kuhn et al., 2018). The optimal number of PLS components for each model was defined by the minimum value of root mean square of standard error (RMSE) obtained after a 10-folds cross validation. The maximum number of latent variables in the model was set to 20. The performance of each regression model was evaluated using RMSE and the coefficient of determination (R^2) of calibration and cross-validation. As the number of samples was much higher than the number of components in the model, RMSE was calculated as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{Nsample}(y_i - \hat{y}_i)^2}{Nsample}}$$

Before chemometric analysis, the centred and scaled spectral dataset was split into three groups according to the randomly chosen sampling dates (R package "cvTools", version 0.3.2; Alfons, 2012). Spectral data from 2 groups were used as the calibration set and the remaining group data as the validation set. This selection procedure was repeated three times in order to assess the variability of performance statistical parameters. Therefore, this validation is a second cross-validation in terms of sampling dates. ATR-MIR train and test sets contained 155 and 77 records respectively. The first NIR train and test sets contained 153 and 67 records. The second NIR train and test sets included 151 and 69 records. The third NIR train and test sets contained 142 and 78 records.

3. Results

3.1. Control milk samples

Table 6-2: Contents of the major milk components (mean ± SD) in the used samples determined by MPA spectrometer (Bruker Optics, Ettlingen). The *P*-values of the paired t test between adulterated and control milk samples are mentioned within brackets.

Adulterant Type	Level (g/L)	Fat (g/100 g)	Protein (g/100 g)	Lactose (g/100 g)	Total solids (g/100 g)
Control	0	3.15±0.78	3.21±0.12	4.88±0.06	12.09±0.85
HRP	40	3.50±0.69 (ns)	5.56±0.16 (***)	4.62±0.10 (***)	14.22±0.73 (***)
	20	3.51±0.71 (ns)	4.36±0.21 (***)	4.76±0.11 (**)	13.21±0.75 (**)
	10	3.31±0.70 (ns)	3.77±0.15 (***)	4.81±0.08 (§)	12.65±0.76 (ns)
	5	3.32±0.56 (ns)	3.50±0.13 (***)	4.83±0.06 (ns)	12.49±0.64 (ns)
	2.5	3.32±0.74 (ns)	3.37±0.12 (ns)	4.83±0.07 (ns)	12.37±0.84 (ns)
HWP	30	3.30±0.70 (ns)	5.38±0.20 (***)	5.36±0.05 (***)	14.10±0.78 (***)
	15	3.31±0.74 (ns)	4.22±0.39 (***)	5.04±0.12 (**)	13.06±0.91 (*)
	7.5	3.26±0.69 (ns)	3.80±0.18 (***)	4.96±0.06 (*)	12.67±0.83 (ns)
	3.75	3.27±0.70 (ns)	3.56±0.23 (**)	4.89±0.10 (ns)	12.45±0.80 (ns)
	1.88	3.32±0.68 (ns)	3.39±0.15 (**)	4.88±0.04 (ns)	12.39±0.80 (ns)
Urea	8	3.31±0.64 (ns)	3.14±0.13 (ns)	4.76±0.07 (***)	11.91±0.74 (ns)
	4	3.64±1.16 (ns)	3.17±0.12 (ns)	4.79±0.08 (*)	12.38±1.16 (ns)
	2	3.60±1.00 (ns)	3.18±0.12 (ns)	4.82±0.08 (§)	12.42±1.03 (ns)
	1	3.48±0.83 (ns)	3.19±0.11 (ns)	4.83±0.08 (ns)	12.36±0.90 (ns)
	0.5	3.41±0.80 (ns)	3.20±0.12 (ns)	4.84±0.06 (ns)	12.31±0.89 (ns)

Table 6-2 continued

Adulterant Type	Level (g/L)	Fat (g/100 g)	Protein (g/100 g)	Lactose (g/100 g)	Total solids (g/100 g)
Water	50	3.07±0.67 (ns)	3.05±0.11 (**)	4.64±0.05 (***)	11.70±0.76 (ns)
	25	3.15±0.69 (ns)	3.11±0.11 (§)	4.76±0.06 (***)	11.93±0.78 (ns)
	12.5	3.18±0.73 (ns)	3.16±0.12 (ns)	4.81±0.07 (*)	12.05±0.82 (ns)
	6.25	3.23±0.79 (ns)	3.19±0.13 (ns)	4.86±0.05 (ns)	12.15±0.88 (ns)
	3.13	3.25±0.75 (ns)	3.19±0.12 (ns)	4.85±0.06 (ns)	12.18±0.83 (ns)
Whey	30	3.47±0.65 (ns)	5.27±0.18 (***)	4.60±0.10 (***)	14.48±0.72 (***)
	15	5.45±0.68 (ns)	6.70±0.14 (***)	7.22±0.08 (***)	13.32±0.76 (**)
	7.5	3.35±0.68 (ns)	3.72±0.12 (***)	4.80±0.07 (*)	12.80±0.79 (§)
	3.75	3.35±0.75 (ns)	3.47±0.12 (**)	4.82±0.06 (§)	12.54±0.85 (ns)
	1.88	3.37±0.78 (ns)	3.33±0.12 (*)	4.84±0.07 (ns)	12.43±0.87 (ns)

Notes: HRP = hydrolysed rice protein; HWP = hydrolysed wheat protein. *P* values indicate the significance of paired *t*-tests of composition values between adulterated samples and control milk. ns = $P > 0.1$; § = $0.05 < P < 0.1$; * = $0.01 < P < 0.05$; ** = $0.001 < P < 0.01$; *** = $P < 0.001$.

Milk samples were collected from 9 different cows in 2 herds. The idea behind this particular sampling was to increase the natural variability of milk composition within the calibration and validation datasets. The mean and SD values for the main milk components are given in Table 2. The most variable milk component from the 9 milk samples used was the fat content (coefficient of variation (CV) = 24.93%) with a minimum and maximum value of 1.66% and 4.84%, respectively. The second most variable component was the total solids content (CV = 7.00%, maximum = 13.74% and minimum = 10.35%). This is due to the variability observed for the milk fat and protein contents. The average value for protein content was 3.21% with a minimum and maximum value of 2.90% and 3.29% (CV = 3.59%) respectively. The quantity of lactose was more stable in the collected samples. The content varied from 4.77 to 4.94% with a mean of 4.88% (CV = 1.29%).

3.2. Composition of adulterated milk samples

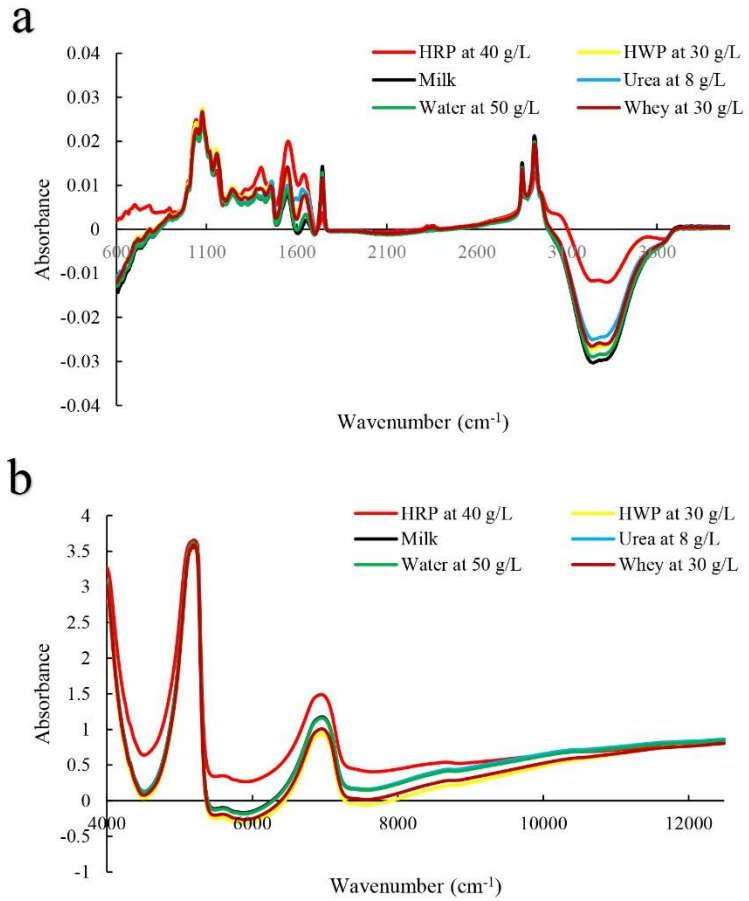


Figure 6-1: Average ATR-MIR (a) and NIR (b) spectra of control milk and samples with targeted concentration of adulterants [hydrolyzed rice protein (HRP), 40 g/L; hydrolyzed wheat protein (HWP) and whey, 30 g/L; urea, 8 g/L; water, 50 g/L].

The composition of adulterated milk samples predicted using the MPA spectrometer is listed in Table 6-2. Compared with control milk, no significant changes in fat predictions were induced by the addition of adulterants to milk. However, remarkable increases in predicted protein values were observed in HRP-, HWP-, and whey-adulterated milk samples. Higher levels of water addition (50 and 25 g/L) caused a lower prediction of protein content in adulterated samples than the protein content measured in control milk. In this study, medium to high levels of HWP added to samples ($\geqslant$ 7.5 g/L) increased the predictions of lactose, while corresponding levels of other adulterants reduced the prediction of lactose content in samples. In addition, high levels of HRP, HWP, and whey in adulterated samples ($\geqslant$ 15 g/L) also resulted

in an increase of the total solids contents in samples, explained by the increase in protein predictions. These results allow us to conclude that the addition of adulterants to milk changed the NIR milk spectra, as the predictions produced by the MPA spectrometer are based on NIR transmittance.

3.3. Spectra characteristic of adulteration

Average ATR-MIR and NIR spectra for control milk and samples spiked with the highest level of adulterant are presented in Figure 6-1a and 6-1b, respectively. As expected, based on the observed milk compositions, spectral changes can be seen between samples.

Major MIR spectral differences between adulterated samples focused on the regions of 600-850, 1 450-1 720, and 3 000-3 600 cm^{-1}. HRP samples showed higher absorbance than other samples. For all samples, two primary peaks centred at 1 645 and 1 550 cm^{-1} were observed. For urea-spiked samples, there were 2 extra minor peaks at 1 596 and 1 630 cm^{-1}. In addition, HWP and whey samples also showed higher absorbance than averaged control milk and samples adulterated with water in the region of 1 450-1 720 cm^{-1}.

All samples showed two prominent NIR bands centred at 6 950 and 5 200 cm^{-1}. HRP samples showed higher absorbance than other samples in NIR regions of 4 190-4 930 and 5 340-8 940 cm^{-1}.

Scores plots for PCA obtained with MIR and NIR spectra did not show any samples clusters according to adulterant type or level, sampling date, or dilution (data not shown).

3.4. Classification model per adulterant type

PLS-DA models were developed to discriminate adulterated milk samples from 9 different control milk samples. The classification performances following the 5 spectral pre-treatments based on ATR-MIR and NIR spectra are summarised in Table 6-3.

For ATR-MIR, the cross-validation accuracy ranged from 0.51 to 0.65. Similar results were obtained for the validation based on the sampling data (i.e., the second cross-validation). Lower values were estimated for the Cohen's kappa parameters. The value ranged from 0.39 to 0.57 for the cross-validation and between 0.40 and 0.50 for the sampling date validation. All of these results suggest a moderate performance of ATR-MIR in classifying samples according to adulterant type. However, the use of spectral pre-treatment provided worse classification performances compared to those without pre-treatment for ATR-MIR.

For NIR, the accuracy obtained based on the sampling date validation was similar to that observed for MIR and ranged from 0.52 to 0.58, although higher values were observed for the 10-fold cross-validation (0.64-0.70). The Cohen's kappa parameter of no pre-treatment data for NIR of cross-validation (0.64) also had values comparable to those observed for ATR-MIR (0.65). However, NIR classification using first

derivative pre-treatment showed better accuracy (0.70), nearly 10% higher than using no pre-treatment data (0.64).

Table 6-3: Mean and SD of accuracy and Cohen's Kappa unweighted parameter obtained from the cross-validation and the sampling date validation for the classification of adulterant type.

				Cross-validation			Validation		
	Pre-treatment		Ncomp	N	Accuracy	Kappa	N	Accuracy	Kappa
ATR-MIR	None	Mean	11	155	0.65	0.57	77	0.60	0.50
		SD	0	0	0.02	0.03	0	0.06	0.07
	1D	Mean	6	155	0.51	0.39	77	0.52	0.40
		SD	2	0	0.09	0.11	0	0.07	0.08
	1D + SNV	Mean	10	155	0.54	0.43	77	0.54	0.43
		SD	3	0	0.03	0.03	0	0.04	0.05
	2D	Mean	5	155	0.58	0.48	77	0.56	0.46
		SD	1	0	0.05	0.06	0	0.06	0.07
	2D + SNV	Mean	6	155	0.52	0.40	77	0.53	0.42
		SD	2	0	0.02	0.03	0	0.06	0.07
NIR	None	Mean	17	149	0.64	0.55	71	0.52	0.40
		SD	2	6	0.04	0.05	6	0.11	0.13
	1D	Mean	9	149	0.70	0.63	71	0.58	0.48
		SD	1	6	0.04	0.06	6	0.10	0.13
	1D + SNV	Mean	11	149	0.64	0.56	71	0.55	0.44
		SD	1	6	0.03	0.04	6	0.10	0.12
	2D	Mean	9	149	0.68	0.60	71	0.57	0.47
		SD	1	6	0.02	0.03	6	0.11	0.13
	2D + SNV	Mean	11	149	0.57	0.47	71	0.52	0.40
		SD	7	6	0.02	0.02	6	0.07	0.09

Notes: ncomp= number of components; N = number of samples; 1D = first derivative of gap 5; 2D = second derivative of gap 5; SNV = standard normal variate.

Although the global accuracy was moderate, based on Figure 6-1a and 6-1b, we can easily assume that the accuracy can depend on the adulterant used. Table 6-4 shows the sensitivity and specificity per adulterant type, calculated based on the sampling date validation. Sensitivity for ATR-MIR ranged from 0.00% for the detection of control milk to 75.56% observed for HWP. For NIR, the sensitivity ranged from 0.00% for milk to 80.95% for water. Specificity ranged from 60.22% to 100% for ATR-MIR

and from 74.44% to 100% for NIR. With milk, whey detection presented the second lowest sensitivity value using ATR-MIR (30.48%). For NIR, it was HWP (56.07%). From Table 6-4, we can easily observe a complementary between NIR and MIR. For instance, NIR gave better results in detecting water and whey, while MIR was better for discriminating HRP, HWP and urea.

From Table 6-3 and 6-4, we can sometimes observe a strong variability in results for sensitivity and specificity from the sampling date validation dataset. This could suggest that the developed classification models are not robust enough. If we pooled all available samples (i.e., 232 for ATR-MIR and 220 for NIR), the calibration accuracy was equal to 0.79 for ATR-MIR and NIR. Cohen's kappa parameters were slightly lower at 0.74 and 0.73 respectively (Table 6-5). Using a 10-fold cross-validation, the classification performances were 0.65 and 0.70 for ATR-MIR and NIR accuracy and 0.57 and 0.63 for ATR-MIR and NIR kappa parameter (in Table 6-3). These values were globally higher than those observed previously using 3 different calibration and validation datasets (in Table 6-3).

Table 6-4: Sensitivity and specificity for the classification of adulterant type using ATR-MIR and NIR spectroscopy from the sampling date validation set.

	Pretreatment		HRP	HWP	Milk	Urea	Water	Whey
Sensitivity (%)								
ATR-MIR	None	Mean	66.67	75.56	0.00	63.01	73.33	30.48
		SD	13.34	10.18	0.00	23.88	20.00	9.07
NIR	1D	Mean	56.30	56.07	0.00	55.40	80.95	55.55
		SD	3.40	14.15	0.00	15.94	21.82	32.88
Specificity (%)								
ATR-MIR	None	Mean	98.92	95.16	100.00	99.46	60.22	96.29
		SD	1.86	5.59	0.00	0.93	11.33	1.82
NIR	1D	Mean	98.30	86.19	100.00	99.37	74.44	89.28
		SD	1.70	11.98	0.00	1.09	7.13	10.34

Notes: 1D = first derivative of gap 5; HRP = hydrolysed rice protein; HWP = hydrolysed wheat protein; milk = control milk; SD = standard deviation.

We observed changes in sensitivity and specificity between the adulterant types. The dilution can also impact the performances of classification. To assess these performances, the NIR and ATR-MIR classifications developed from the full dataset were used in order to have a high enough number of samples per dilution percentage. The results are summarised in Table 6-5. The values obtained for the calibration accuracy and Cohen's kappa parameters increased with the dilution percentages for both NIR and ATR-MIR. This was expected as there is a smaller spectral fingerprint for the adulterant present in the milk spectra when the level of this adulterant is low.

Table 6-6 shows the changes of sensitivity per adulterant type as a function of the dilution percentage. Based on these results from ATR-MIR spectroscopy, a perfect discrimination of HRP and HWP in milk was possible from 6.25% v/v. From 12.5% v/v, it was possible to discriminate urea. Perfect discriminations of water and whey were not possible using ATR-MIR. The use of NIR was less interesting for discriminating HRP, HWP and urea compared to ATR-MIR. However, NIR allowed a perfect discrimination of whey from 12.5% v/v. The discrimination of water adulteration using NIR was largely better than that observed using ATR-MIR (between 88.89% to 100% for NIR and between 66.67% and 77.78% for ATR-MIR).

3.5. Prediction of adulterant level

Performances for adulterant level quantification using ATR-MIR and NIR with different pre-treatments are listed in Table 6-7. Based on these results, as observed for the classification performances of adulterant type, the best performances were observed without pre-treatment for MIR and with a first derivative pre-treatment for NIR. Therefore, these pre-treatments were used for further analysis.

The global quantification of adulterant level provided a cross-validation and sampling date validation R^2 of 0.45 and 0.44 for ATR-MIR, and 0.70 and 0.29 for first derivative NIR, respectively (Table 6-7). Water was present in the dataset used. As this adulterant is not a protein source, this led to some noise in the equation. If the water is left out from the datasets, the prediction performances greatly increased (Table 6-8). For ATR-MIR, cross-validation and date sampling validation R^2 reached 0.88 and 0.87. This is an increase of 95.6% and 97.7%, respectively, compared to the equation constructed with water-adulterated samples. For NIR, these values were equal to 0.99 and 0.97, respectively. This represents an increase of 41.4% and 234.4% compared to the equation built from datasets containing water-adulterated samples. This increase of sampling date validation of NIR was higher than the increase observed for ATR-MIR. This may be related to the higher sensitivity of NIR to water.

Prediction equations were built per adulterant type to observe if the performance of prediction can be improved and to find the less well-predicted adulterant level (Table 6-8). As observed for the classifications of adulterant type, ATR-MIR gave the good ability to predict the level of HRP-, HWP- and urea-spiked milk. The quantification of HWP level by ATR-MIR (validation R^2 = 0.95, RMSE = 3.80 g/L) was better than prediction by NIR (validation R^2 = 0.88, RMSE = 3.70 g/L). The content of whey in adulterated samples was better predicted using NIR (validation R^2 = 0.97, RMSE = 2.10 g/L). The prediction of water content in diluted milk was also better using NIR but the performance was weak (validation R^2 = 0.41 with a RMSE = 21.19 g/L of milk).

4. Discussion

4.1. Spectral characteristics

Table 6-5: Accuracy and Cohen's kappa parameter for classifications based on adulterant type following the dilution percentage. This classification used all available samples.

		Calibration		
		N	Accuracy	Kappa
ATR-MIR	All samples	232	0.79	0.74
	3.13	45	0.58	0.47
	6.25	45	0.82	0.78
	12.5	45	0.87	0.83
	25	44	0.91	0.89
	50	44	0.93	0.91
First derivative NIR	All samples	220	0.79	0.73
	3.13	44	0.43	0.29
	6.25	42	0.74	0.67
	12.5	42	0.98	0.97
	25	42	0.98	0.97
	50	41	1.00	1.00

Notes: MIR = mid-infrared spectroscopy; NIR = near-infrared spectroscopy

The changed composition values (except fat; Table 6-2) between control and adulterated samples observed between major milk components involved differences in the spectral information. The ATR-MIR spectra (in Figure 6-1a) of control milk was similar to a previous study (Aernouts et al., 2011). The sample adulterated with urea showed different peaks from the classical two peaks of Amide I (1 645 cm^{-1}) and II (1 550 cm^{-1}) observed in other samples; a minor peak (1 596 cm^{-1}) appeared between these two peaks, and an extra peak (1 630 cm^{-1}) before the Amide I peak. These two additional peaks involved the C=O absorption of urea (Santos et al., 2013b) and structural modification of milk protein caused by the addition of urea (Santos et al., 2013a), respectively. Higher absorbance of ATR-MIR spectra presented by HRP in the region of 3 000-3 600 cm^{-1} was related to O-H stretching of water absorption (Aernouts et al., 2011). In addition, higher absorption of samples with HRP at 1 400 cm^{-1} was observed and can be associated with Asp and Glu in hydrolysed plant protein (Barth, 2007). Major MIR spectral differences between adulterated samples focused on the region of 1 450-1 720 cm^{-1} and were associated with the characteristic absorption of Amide I and II of milk protein. High absorbance of samples with HRP, medium absorbance of samples with HWP and whey, and low absorbance of control samples and samples adulterated with water, corresponded to the protein contents levels shown in Table 6-2.

The NIR spectra of samples obtained in the current study (Figure 6-1b) were comparable to those presented in other studies (Santos et al., 2013b). Higher

absorbance of samples with HRP at 6 950 cm^{-1} than other samples was associated with overtones of water absorption in the NIR region (Laporte and Paquin, 1999). The region located between 3 996 and 8 000 cm^{-1} is correlated to water absorption (Laporte and Paquin, 1999).

4.2. Classification of adulterant type

In this study, both ATR-MIR and NIR spectroscopy were not be able to separate control milks from adulterated samples (Table 6-4). For comparison, a similar study showed that normalised and second derivative ATR-MIR could separate control milk from adulterated samples but processed NIR could not, for bovine milk adulterated with water, whey, hydrogen peroxide, synthetic milk, synthetic urine, and urea (Santos et al., 2013b). This inability could be explained by a larger variability of milk composition in this study, which was collected from individual cows and not commercial skimmed milk. However, by combining prediction of adulterant level and classification results, it could be possible to make this separation. This hypothesis must be tested using a larger number of control milk samples.

Table 6-6: Sensitivity for the classification of adulterant type in function of the dilution and adulterant type from the sampling date validation set.

	Dilution (% v/v)	N	Sensitivity				
			HRP	HWP	Urea	Water	Whey
ATR-MIR	3.13	45	55.56	66.67	44.44	66.67	55.56
	6.25	45	100.00	100.00	88.89	66.67	55.56
	12.5	45	100.00	100.00	100.00	66.67	66.67
	25	44	100.00	100.00	100.00	66.67	87.50
	50	44	100.00	100.00	100.00	77.78	88.89
	All	232	91.11	93.33	86.36	68.89	70.45
First derivative NIR	3.13	44	12.50	44.44	33.33	88.89	33.33
	6.25	42	37.50	87.50	55.56	100.00	88.89
	12.5	42	100.00	100.00	100.00	88.89	100.00
	25	42	100.00	100.00	100.00	88.89	100.00
	50	41	100.00	100.00	100.00	100.00	100.00
	All	220	66.67	85.71	77.27	93.18	84.44

Notes: HRP = hydrolysed rice protein; HWP = hydrolysed wheat protein; milk = control milk; ATR-MIR = attenuated total reflectance mid-infrared; NIR = near-infrared

A better sensitivity was observed for HRP-adulterated samples using ATR-MIR (91.11%) than NIR (66.67%) using the entire dataset (Table 6-6), and this difference between NIR (56.30%) and ATR-MIR (66.67%) were also obtained in the validation

based on sampling date (Table 6-4). The specificity was equal to 98.92% for ATR-MIR (Table 6-4). Compared to NIR, ATR-MIR also allowed a better classification for a low content of HRP (6.25% v/v; Table 6-6). Similar conclusions to HRP can be formulated for HWP adulteration. Indeed, HWP spiked milk samples were better classified using ATR-MIR spectroscopy with a sensitivity of 75.56% for validation by sampling date (Table 6-4) and 93.33% for the entire dataset (Table 6-6). The observed specificity of the sampling date validation set was equal to 95.16%. Compared to NIR, ATR-MIR also allowed a better classification for a low content of HWP (6.25% v/v). A previous SDS-PAGE study conducted by our team (Yang et al., 2019) showed that the HRP and HWP used in the current study contained small peptides (less than 11 kD for HRP and 11-17 kD for HWP). Smaller peptides fractions indicate a higher degree of hydrolysis of protein components, this may explain the higher absorbance HRP exhibited in the water absorption associated region in NIR and ATR-MIR. Small peptides with high solubility may also explain the better performance of ATR-MIR classification for these two adulterants.

ATR-MIR spectroscopy also allowed a better classification compared to NIR for urea-adulterated samples, with higher sensitivity (63.01% vs. 55.40% in Table 6-4, 86.36% vs. 77.24% in Table 6-6). The best performance was observed for samples adulterated with urea contents equal to at least 12.5% v/v (2 g/L). In other study, when combined with SIMCA, ATR-MIR successfully separated urea < 900 ppm (approximately 0.9 g/L), and urea > 900 ppm, with a classification efficiency of 80.0-97.8% (Jha et al., 2015). Coupled with improved support vector machine (SVM) and K nearest neighbors (KNN), SNV pre-treated NIR spectra would classify samples with urea with validation correct ratios of 86.5-95.3% (Zhang et al., 2014).

Water-spiked milk samples were better classified using NIR than using ATR-MIR (with sensitivities of 80.95% vs. 73.33% in Table 6-4, and of 93.18% vs. 68.89% in Table 6-6). The ATR-MIR results obtained in this study were lower compared to those obtained by Botelho et al. (2015). From PLS-DA models combined with ATR-MIR, they were able to detect the presence of adulterated water (0.05-1 g/L) in milk, with sensitivity of 93.8% and specificity of 88.5% for validation (Botelho et al., 2015). Samples adulterated with water (150 g/L) were not efficiently differentiated from unadulterated samples by Gondim et al. (2017) due to low specificity (56.7%) by using a SIMCA model with ATR-MIR spectra processed using multiplicative scatter correction (MSC).

In this study, PLS-DA models allowed a better discrimination of whey-spiked milk samples using NIR spectral data than the same samples using ATR-MIR data (with sensitivities of 55.55% vs. 30.48% in Table 6-4, and of 84.44% vs. 70.45% in Table 6-6). Discriminant PLS using NIR (original, MSC, or MSC + second derivative processing) recognised milk samples adulterated with whey (2.15-48.4% g/v, around 21.5-484 g/L) and water (1-97% v/v, around 10-970 g/L), with correct classification of 88.8-100% in validation (Kasemsumran et al., 2007).

4.3. Quantification of adulterant level

Table 6-7: Prediction of the adulterant level in milk from ATR-MIR and NIR spectroscopy.

				Calibration		Cross-validation		Validation	
	Pre-treatment		ncomp	R^2c	RMSE	R^2cv	RMSE	R^2p	RMSEP
ATR-MIR	None	Mean	6	0.60	8.17	0.45	9.52	0.44	9.98
		SD	1	0.03	0.26	0.02	0.06	0.06	0.73
	1D	Mean	2	0.49	9.20	0.37	10.33	0.33	11.00
		SD	1	0.13	1.25	0.04	0.38	0.06	0.66
	1D + SVN	Mean	3	0.61	8.12	0.34	10.60	0.36	10.45
		SD	0	0.04	0.40	0.06	0.42	0.09	0.78
	2D	Mean	2	0.49	9.22	0.37	10.33	0.34	10.97
		SD	0	0.03	0.25	0.05	0.35	0.02	0.47
	2D + SNV	Mean	2	0.53	8.83	0.32	10.69	0.26	11.27
		SD	1	0.12	1.23	0.09	0.40	0.07	0.63
NIR	None	Mean	5	0.39	10.49	0.32	10.47	0.26	11.65
		SD	1	0.02	0.04	0.01	0.02	0.04	0.30
	1D	Mean	14	0.45	10.10	0.70	6.09	0.29	14.29
		SD	9	0.07	0.61	0.38	4.13	0.04	2.78
	1D + SVN	Mean	5	0.38	10.45	0.34	10.35	0.25	11.81
		SD	1	0.04	0.26	0.04	0.22	0.06	0.66
	2D	Mean	2	0.45	10.38	0.30	10.65	0.25	11.12
		SD	0	0.04	0.25	0.02	0.14	0.06	0.35
	2D + SNV	Mean	3	0.39	10.62	0.35	10.25	0.21	12.05
		SD	1	0.03	0.12	0.06	0.47	0.02	0.50

Notes: RMSE = root mean square error; R^2 = coefficient of determination; SD = standard deviation; ncomp = number of components; N = number of samples; 1D = first derivative of gap 5; 2D = second derivative of gap 5; SNV = standard normal variate.

Table 6-8: Prediction of the adulterant level in milk per adulterant type from ATR-MIR and NIR spectroscopy.

				Calibration			Cross-validation		Validation		
			ncomp	N	R^2c	RMSEC	R^2cv	RMSEcv	N	R^2p	RMSEP
ATR-MIR	Without water	Mean	8	125	0.94	2.66	0.88	3.85	62	0.87	4.10
		SD	0	0	0.01	0.27	0.05	0.69	0	0.05	0.85
	HRP	Mean	7	30	0.99	1.09	0.97	2.33	15	0.97	2.55
		SD	1	0	0.01	0.61	0.01	0.40	0	0.01	0.57
	HWP	Mean	10	30	1.00	0.34	0.93	2.30	15	0.95	3.80
		SD	3	0	0.00	0.52	0.04	0.09	0	0.03	2.30
	Urea	Mean	9	29	1.00	0.07	0.97	0.46	15	0.98	0.41
		SD	2	1	0.00	0.07	0.01	0.06	1	0.01	0.10
	Water	Mean	6	30	0.65	8.68	0.69	14.78	15	0.33	15.91
		SD	3	0	0.30	6.12	0.05	1.88	0	0.16	3.23
	Whey	Mean	8	30	0.97	1.51	0.74	5.64	14	0.40	9.10
		SD	2	1	0.02	0.92	0.18	1.89	1	0.28	1.85
First derivative NIR	Without water	Mean	15	119	0.99	0.66	0.99	0.93	57	0.97	1.95
		SD	7	5	0.01	0.75	0.02	0.48	5	0.02	0.23
	HRP	Mean	11	25	1.00	0.05	0.99	0.67	11	1.00	1.47
		SD	1	3	0.00	0.02	0.00	0.00	3	0.00	0.15
	HWP	Mean	11	29	0.98	0.89	0.99	1.33	13	0.88	3.70
		SD	8	2	0.04	1.47	0.02	1.19	2	0.07	1.17
	Urea	Mean	15	30	1.00	0.00	0.99	0.17	14	0.97	0.49
		SD	2	1	0.00	0.00	0.01	0.07	1	0.03	0.30
	Water	Mean	13	30	1.00	0.52	0.82	7.25	14	0.41	21.19
		SD	4	1	0.00	0.45	0.06	1.20	1	0.19	9.76
	Whey	Mean	7	30	1.00	0.37	1.00	0.81	15	0.97	2.10
		SD	2	0	0.00	0.16	0.00	0.07	0	0.02	0.23

Notes: HRP = hydrolysed rice protein; HWP = hydrolysed wheat protein; RMSE = root mean square error; R^2 = coefficient of determination; SD = standard deviation; ncomp = number of components; N = number of samples;

More factors were likely to be included in the final models using NIR than those using ATR-MIR (Tables 6-7 and 6-8), which may suggest that the overlapped overtone and combination bands in NIR spectra need more dimensional analysis to extract effective information. Globally similar prediction performances were observed with NIR and ATR-MIR, although some differences appears (Table 6-8). The global content of protein adulterant (i.e., without water-adulterated samples) was better

predicted using NIR than ATR-MIR (validation R^2p = 0.97 vs. 0.87 with a RMSEP of 1.95 vs. 4.11 g/L), but the main difference appeared for the prediction of whey levels in spiked milk. Validation R^2 for the whey level was equal to 0.40 for ATR-MIR and 0.97 for NIR, with a validation RMSE of 9.10 and 2.10 g/L, respectively. In other studies, combined with PLS regression, well performed quantifications of whey level in adulterated samples using ATR-MIR (validation R^2 = 0.96-0.98, SEP = 1.18-2.33 g/L) (Santos et al., 2013a, Santos et al., 2013b) and NIR (validation R^2 = 0.999, RMSE = 0.244-0.802) (Kasemsumran et al., 2007) have been reported.

For NIR, a slightly better prediction of HRP levels (validation R^2p = 1.00, RMSEP = 1.47 g/L) was obtained for HWP levels (R^2p = 0.88, SEP = 3.70 g/L). Similar performances of ATR-MIR prediction for HRP (R^2p = 0.97, RMSEP = 2.55 g/L) and HWP (R^2p = 0.95, RMSEP = 3.80 g/L) levels in adulterated samples were obtained by PLS regression models. Prediction accuracy of plant protein level in this study is better than the prediction (R^2p = 0.98, RMSEP = 43.6 g/L) for soy milk levels in cow-buffalo milk found using multi-linear regressions with ATR-MIR spectra (Jaiswal et al., 2015).

Both spectroscopies showed excellent performance in predicting urea levels in adulterated milk, and this is close to the accuracy of ATR-MIR prediction of urea levels using spectra data for dried sample film with a PLS regression model (validation R^2 = 0.98, SEP = 0.23-0.30 g/L) (Santos et al., 2013a, Santos et al., 2013b), and better than ATR-MIR prediction (validation R^2 = 0.88, RMSE = 0.24 g/L) of urea levels in fluid milk using PLS regression models (Jha et al., 2015) in previous reports.

In our study, ATR-MIR and NIR quantified water levels in samples with poor accuracy (validation R^2 around 0.40, Table 8). However, another published paper has demonstrated the ability of NIR to predict on water adulteration (1.00-97.00% v/v, around 10-970 g/L) in milk, with validation R^2 of 0.992-0.997 and RMSE of 2.159-3.702 (Kasemsumran et al., 2007). Finally, both spectroscopies in this study showed their favorable performance (validation R^2p = 0.87-0.98, SEP = 1.95-4.10 g/L) for total protein adulterants levels prediction in samples.

Consequently, some differences in terms of quantitative or qualitative performances appeared in comparison to previously published papers. A potential explanation could be the use of individual cow milks as control milk. In the other studies, the control milks used are often pooled commercial (Santos et al., 2013b) or raw milks (Jaiswal et al., 2015), which leads a decrease in natural protein variability. The increase of natural protein as used in this study can interact in some infrared regions with the adulteration-based proteins, suggesting a lower ability to detect low adulterated samples.

5. Conclusions

This study explores the potential of ATR-MIR and NIR spectroscopy to detect adulteration of milk with protein or nitrogen-rich compounds, especially with hydrolysed plant protein. The classification model could not differentiate adulterated samples from control milk, probably due to the large variability in cow milk

composition. Meanwhile, an obvious effect of dilution on classification performance was observed. Low diluted samples (dilution percentage ⩾ 12.5%) with HRP, HWP, and urea were classified by both spectroscopies. Compared with MIR, NIR spectra provided a more accurate quantification of HRP and whey, while MIR prediction of HWP and urea levels was better than NIR prediction. Moreover, both spectroscopies showed their good properties (validation R^2 = 0.87-0.98, RMSE = 2.04-4.11 g/L) on the prediction of protein adulterants. However, a large variability of performance results was observed between training and validation sets suggesting the need to increase the number of individual cow milk samples used.

6. Acknowledgements

We would like to express our sincere thanks to Anne-Catherine Dalcq for her help with milk sampling, as well as Nicolas Crasset, Quentin Arnould, Bernard Lecler, and Ouissam Abbas for their help with spectra collection and data export. We are grateful for the instrumental support from the Walloon Agricultural Research Centre (CRA-W).

7. References

Aernouts, B., E. Polshin, W. Saeys, and J. Lammertyn. 2011. Mid-infrared spectrometry of milk for dairy metabolomics: a comparison of two sampling techniques and effect of homogenization. Anal Chim Acta 705(1-2):88-97.

Balabin, R. M. and S. V. Smirnov. 2011. Melamine detection by mid- and near-infrared (MIR/NIR) spectroscopy: A quick and sensitive method for dairy products analysis including liquid milk, infant formula, and milk powder. Talanta 85(1):562-568.

Barth, A. 2007. Infrared spectroscopy of proteins. Biochimica et Biophysica Acta (BBA) - Bioenergetics 1767(9):1073-1101.

Botelho, B. G., N. Reis, L. S. Oliveira, and M. M. Sena. 2015. Development and analytical validation of a screening method for simultaneous detection of five adulterants in raw milk using mid-infrared spectroscopy and PLS-DA. Food Chem 181:31-37.

Gondim, C. d. S., R. G. Junqueira, S. V. C. d. Souza, I. Ruisánchez, and M. P. Callao. 2017. Detection of several common adulterants in raw milk by MID-infrared spectroscopy and one-class and multi-class multivariate strategies. Food Chem 230:68-75.

Handford, C. E., K. Campbell, and C. T. Elliott. 2016. Impacts of milk fraud on food safety and nutrition with special emphasis on developing countries. Compr Rev Food Sci F 15(1):130-142.

Ho, M. H.-K., W. H.-S. Wong, and C. Chang. 2014. Clinical spectrum of food allergies: a comprehensive review. Clinical reviews in allergy & immunology 46(3):225-240.

Jaiswal, P., S. N. Jha, A. Borah, A. Gautam, M. K. Grewal, and G. Jindal. 2015. Detection and quantification of soymilk in cow-buffalo milk using Attenuated Total

Reflectance Fourier Transform Infrared spectroscopy (ATR-FTIR). Food Chem 168:41-47.
Jha, S. N., P. Jaiswal, A. Borah, A. K. Gautam, and N. Srivastava. 2015. Detection and quantification of urea in milk using attenuated total reflectance-fourier transform infrared spectroscopy. Food and Bioprocess Technology 8(4):926-933.
Kasemsumran, S., W. Thanapase, and A. Kiatsoonthon. 2007. Feasibility of near-infrared spectroscopy to detect and to quantify adulterants in cow milk. Anal Sci 23(7):907-910.
Laporte, M.-F. and P. Paquin. 1999. Near-infrared analysis of fat, protein, and casein in cow's milk. J Agr Food Chem 47(7):2600-2605.
Lohumi, S., S. Lee, H. Lee, and B.-K. Cho. 2015. A review of vibrational spectroscopic techniques for the detection of food authenticity and adulteration. Trends Food Sci Tech 46(1):85-98.
Luykx, D. M. A. M. and S. M. van Ruth. 2008. An overview of analytical methods for determining the geographical origin of food products. Food Chem 107(2):897-911.
Maraboli, A., T. M. P. Cattaneo, and R. Giangiacomo. 2002. Detection of vegetable proteins from soy, pea and wheat isolates in milk powder by near infrared spectroscopy. J near Infrared Spec 10(1):63-69.
Nascimento, C. F., P. M. Santos, E. R. Pereira, and F. R. P. Rocha. 2017. Recent advances on determination of milk adulterants. Food Chem 221:1232-1244.
Poonia, A., A. Jha, R. Sharma, H. B. Singh, A. K. Rai, and N. Sharma. 2017. Detection of adulteration in milk: A review. International Journal of Dairy Technology 70(1):23-42.
Renny, E., D. Daniel, A. Krastanov, C. Zachariah, and R. Elizabeth. 2005. Enzyme based sensor for detection of urea in milk. Biotechnology & Biotechnological Equipment 19(2):198-201.
Rodriguez-Saona, L. E. and M. E. Allendorf. 2011. Use of FTIR for rapid authentication and detection of adulteration of food. Annu Rev Food Sci T 2:467-483.
Rodriguez-Saona, L. E., F. S. Fry, and E. M. Calvey. 2000. Use of Fourier transform near-infrared reflectance spectroscopy for rapid quantification of castor bean meal in a selection of flour-based products. J Agr Food Chem 48(11):5169-5177.
Santos, P. M., E. R. Pereira-Filho, and L. E. Rodriguez-Saona. 2013a. Rapid detection and quantification of milk adulteration using infrared microspectroscopy and chemometrics analysis. Food Chem 138(1):19-24.
Santos, P. M., E. R. Pereira, and L. E. Rodriguez-Saona. 2013b. Application of hand-held and portable infrared spectrometers in bovine milk analysis. J Agr Food Chem 61(6):1205-1211.
Wang, L., F. S. C. Lee, X. Wang, and Y. He. 2006. Feasibility study of quantifying and discriminating soybean oil adulteration in camellia oils by attenuated total reflectance MIR and fiber optic diffuse reflectance NIR. Food Chem 95(3):529-536.
Whitfield, R. G., M. E. Gerger, and R. L. Sharp. 1987. Near-Infrared Spectrum Qualification via Mahalanobis Distance Determination. Appl. Spectrosc. 41(7):1204-1213.
Yang, J., N. Zheng, H. Soyeurt, and J. Wang. Detection of plant protein in adulterated milk using nontargeted nano-high-performance liquid chromatography–tandem mass

spectroscopy combined with principal component analysis. Food Science and Nutrition.
Zhang, L. G., X. Zhang, L. J. Ni, Z. B. Xue, X. Gu, and S. X. Huang. 2014. Rapid identification of adulterated cow milk by non-linear pattern recognition methods based on near infrared spectroscopy. Food Chem 145:342-348.

7

General discussion, conclusions and perspectives

Previous chapters have depicted the detection of foreign protein, especially (hydrolyzed) plant protein in milk protein adulteration using two-dimensional electrophoresis, high performance liquid chromatography tandem mass spectrum, mid- and near- infrared spectroscopy. This chapter presented these methodologies in detail, discussed the procedure of these methods, compared different methods based on their advantages and limitations, and provided suggestions about the potential improvements in future studies.

1. Comparison of detection methods

In order to guarantee milk quality, series of methodologies are used to detect foreign protein in dairy products. Classical routine analytical tests for protein measurements, such as Kjedahl and Dumas, are non-specific and do not distinguish foreign nitrogen from milk protein nitrogen (Moore et al., 2010). This represents a risk factor for the incidence of economic motivate adulteration (Everstine et al., 2013). To reduce the reliance on detection of total protein nitrogen, the United States Pharmacopeia Convention (USP) has encouraged the development of methodologies to detect economic motivate adulteration (Moore et al., 2010). Due to the uncertainty of the presence of adulterant in milk, untargeted detection would be preferable in practice (Lu et al., 2017); and combined with chemometrics, they allow to differentiate adulterated food against control food (Esslinger et al., 2014). In this thesis, two-dimensional electrophoresis (2- DE), high performance liquid chromatography (HPLC) tandem mass spectroscopy (MS/MS), and near- and mid- infrared (NIR/MIR) spectroscopy were used to identify foreign protein or nitrogen compounds spiked in milk. Details for these methodologies were listed in Table 7-1.

Table 7-1: Comparison of methodology used in this thesis

Items	2-DE	HPLC-MS/MS	NIR/MIR
Instrument	The PROTEAN® i12™ first-dimension isoelectric focusing (IEF) system, Large format PROTEAN® II xi cells, GS800 calibrated densitometer, from BIO-RAD	Nano- HPLC coupled with Q Exactive HF Mass Spectrometry, from Thermo Scientific	MPA FT-NIR spectrometer/ Tensor 27 FT-MIR spectrometer, from Bruker Optics
Price of instrument (RMB, Yuan)	350 000	3 000 000	230 000/450 000
Procedures of each methodology	Fat removal → protein concentration test → isoelectric focusing → reduction and alkylation → SDS-PAGE → staining → destaining → gel scanning → data analysis	Centrifugation → lysis → centrifugation → protein concentration test → reduction and alkylation → trypsin digestion → desalting and drying → HPLC separation → tandem MS detection → data analysis	(Fat removal in Chapter 5) → spectral acquisition → data export → data analysis
Samples preparation	2 steps before isoelectric focusing	7 steps before HPLC separation	Little or no separation
Manual steps	7 steps before data analysis	7 steps before HPLC separation	Steps before spectral acquisition

(Table 7-1 continued)

Items	2-DE	HPLC-MS/MS	NIR/MIR
Toxic regent involved in operation	Urea, acrylamide, thiourea, CHAPS, DTT, SDS, iodoacetamide	Urea, DTT, iodoacetamide, acetonitrile, formic acid	Chloroform for fat removal (Chapter 5) or none (Chapter 6)
Data analysis	Protein spots matching	Raw files processing, data transformation and normalization, PCA	Data transformation and normalization, removing of outliers, PCA, dataset splitting, PLS-DA, PLS regression, prediction of samples in validation set
Special software used	PDQuest	Maxquant, Unscrambler	R software
Duration of measurements	At least 2 days	2 hours	2-4 min
Coefficient of variation (CV, %)	3.2-55.7% for optical intensities of protein spots, with median value of 12.7%.	Median CV values of remarkable peptides intensities ranged from 0.684 to 1.392%	Prediction on milk composition varied, CV for prediction of fat, protein, protein, lactose, and total solids, were 24.93, 3.58, 1.29, and 7.00, respectively.
Detection limit (%, percentages of foreign protein in total protein)	4%	0.5-2%	8%

1.1. 2-DE

Electrophoresis is a powerful method for protein separation. Polyacrylamide gel electrophoresis (PAGE) and its modification are useful to identify and detect the protein adulteration (Moore et al., 2010). For example, 2-DE separates protein fraction with high resolution according to isoelectric point and molecular weight. In Chapter 2, the identified protein fractions (above 35 kDa) have higher molecular weight than dominant milk protein (casein and whey protein, below 35 kDa), while these marked proteins shared the same isoelectric points region (pH 4.9-6.2) with milk protein. Therefore, 2-DE could separate those foreign proteins from milk protein based on their molecular weight, and the results were easy to identify the remarkable adulterants protein. However, the application of this electrophoresis is limited due to the use of special regent and its time-consuming and labour-intensive procedure for sample analysis. As presented in Table 7-1, some used regents (for instance,

dithiothreitol, iodoacetamide) were toxic and particular for molecular biology experiment, not common for routine milk laboratory. The whole procedure running (sample preparation and electrophoresis) costed at least 48 hours, and there is not an automatic equipment for this method currently. Technicians have to carry on isoelectric electrophoresis and SDS-PAGE separately and successively. The consecutive manual operation steps accumulated manual measurement error in the final results. So, high variation of optical intensities of remarkable protein points (CV ranged 3.2-55.7%) was observed in Chapter 2 and really influenced the discriminability of this method, which resulted in poor reproducibility of this method. In addition, the detection limits of this method on adulterated plant protein (4%, in Chapter 2) or milk protein (1%; (Yang et al., 2014) were higher than immunological assays (0.5%; Table 1-2). However, the low detection limit and the poor reproducibility of 2-DE observed in this thesis could be likely improved by the application of two-dimensional differential in-gel electrophoresis (2D-DIGE) and cyanine dyes (Issaq and Veenstra, 2008) in further study.

1.2. HPLC-MS/MS

Compared with electrophoresis, LC-MS is more often used in component identification and adulteration detection in food commodities (Lohumi et al., 2015). HPLC completes the separation of compounds according to three primary characteristics: polarity, electric charge, and molecular size (Danezis et al., 2016). Due to the complexity of samples and high demand of foodstuff analysis, high resolution LC, such as HPLC-MS/MS, was employed to identify food authentication.

Compared with 2-DE, HPLC-MS/MS had a more complicated sample preparation, including lysis, reduction and alkylation, trypsin digestion, as presented in Table 7-1, and also involved expensive (trypsin) and toxic regents (DTT and iodoacetamide). This process is a little labor-intensive, there are 7 steps to complete by technicians before sample boarding, and these toxic regents are not so friendly to healthy of operators and environment. Whereas, unlike 2-DE, automated sampling and measurements of HPLC-MS/MS released technicians from operation, and this automated operation minimized the influence of manual operation on the final results, therefore, good reproducibility was observed on measurements of HPLC, with CV of remarkable peptides intensities ranged 0.68-1.39%. The well-developed procedure of HPLC-MS/MS ask operators not only to know sample preparation well, but also to set instrument parameters and process the results using special software. A qualified operator needs too much training about methodology, instrument, and software before real operation. The whole HPLC-MS/MS running for one sample measurement last 2 hours, and this maybe a little long for continuous samples detection. In recent studies, fingerprints of flow injection mass spectrometry (FIMS) with chemometrics allowed the detection of soybean, pea, and whey protein in milk at the level of 0.5%. The prediction accuracy of super vector model for adulterated and unadulterated samples were 92.86 and 86.75% respectively. More important, FIMS could complete the analysis of per sample within 1 min, without trypsin hydrolysis and chromatographic separation (Du et al., 2018).

1.3. Infrared spectroscopy

Due to their time-consuming and labour-intensive characteristic, 2-DE and HPLC-MS/MS do not allow the analysis of a large number of samples as asked for a routine monitoring of the protein adulteration in milk. Therefore, there is a necessity to develop a rapid and cheap method to detect protein adulteration in milk. The MIR spectrometry is largely used in milk laboratories for the routine prediction of the main milk components such as fat, protein, lactose and urea. Although NIR is less largely used by milk laboratories compared to MIR, this technic can be interesting to detect protein adulteration with relatively low water absorption (Rodriguez-Saona and Allendorf, 2011). Both MIR and NIR were used to detect foreign protein adulterated in raw milk in Chapter 6 of this thesis.

As presented in Table 7-1, with little or without sample preparation, MIR or NIR cost only 2-4 min for one sample running, less than 2-DE and HPLC-MS/MS, and almost no regent involved in measurements, so friendly to technicians and environment. Most of work for this methodology concentrated on data analysis, such as data transformation, model development, and model validation. Corresponding chemometrics may be difficult for some routine users. Another potential limitation for IR is low sensibility to low concentration of adulterants, as presented in Character 6. Some instrument and sample factors affected model performance are discussed as followed.

Different measurement modes affect the performance of IR method. NIR used in Chapter 6 was in transmission mode and sensitive to quality changes (Lohumi et al., 2015). Different measurement modes, such as ATR, diffuse reflectance, high throughput transmission (HTT), and transmission cell, are available for MIR spectroscopy, and ATR-MIR used in Chapter 5 and 6, is the most widely used methods for food authenticity and adulteration (Lohumi et al., 2015). In addition, development and application of ATR is expected to reduce intense water absorption for MIR (Ellis et al., 2012).

Sample composition variation is a crucial factor for the efficiency of spectroscopy model. First, major spectral difference between adulterated samples are related with adulterant types and protein contents, such as absorption of adulterants corresponding to protein contents of samples, as well as additional peaks of samples with urea in Figure 6-1. Second, the non-targeted compounds (such as fat and water) in samples were also associated with model performance. In Chapter 5, skimmed samples were prepared, and ATR-MIR combined with PLS-DA would discriminate almost all adulterated samples according to adulterants types (Table 5-2), detection limit for HSP and HWP were 1.875 g/L (Table 5-3); however, there was only one mixed commercial milk as control sample, which means the effect of fat absorption and sample variation on spectral data were not considered in this part. When 9 raw milks were designed as control milks (Chapter 6), this methodology was unable to separate adulterated samples from authentic milk. Variance of raw milk composition (CV ranged from 1.29% for lactose to 24.93% for fat) was likely the main reason for the poor performance observed for the classification model on HWP and HRP on 3.13% in Table 6-6. Sample inhomogeneity, particle size, and preferential adsorption of fat would

contribute to a lower accuracy of ATR-MIR (Karoui et al., 2006). MIR spectroscopy achieved a better prediction on fat for homogenized samples ($R^2 > 0.92$) than for raw milks ($R^2 < 0.70$) (Aernouts et al., 2011).

Model optimization and validation are also important for the development of spectroscopic methods. In Chapter 6, raw ATR-MIR and first derivative NIR data have shown their better performance than other data transformation. Cross-validation after dataset splitting were often applied in other report (Santos et al., 2013) on spectroscopic methods, and only parameters for one model were presented. In order to observe the effect of raw milk variation on model performance, a second validation in terms of sampling date was brought into Chapter 6. In contrast, absence of external validation in Chapter 5 provided not generalized results.

1.4. Comparison of test methods

Table 7-2: Advantages and limitations of test methods

Methods	Advantages	Limitations
2-DE	High resolution, easy data processing	Toxic regent, time-consuming, labour-intensive, varied optical intensities, manual steps
HPLC-MS/MS	High accuracy, precision, and stability, low detection limit, considered as official methods	Expensive instrument, skilled operators, special software to process data, toxic regent, sample preparation
MIR/NIR	Rapid, non-invasive, little sample preparation, rich information, no pollution, easy to operate	Non-targeted component interference, low accuracy, low sensibility, spectral technology (mode), model validation

In a previous literature review on the technology of food authentication, Danezis et al. (2016) found that, the most extensive application of molecular techniques, such as 2-DE, was on determination of species and botanical origin, while LC and IR spectroscopy mainly on geographic traceability and adulteration. MS has become a frontline detection method on food adulteration with high sensitivity, selectivity, and throughput, especially coupled with LC (Danezis et al., 2016). Based on details about the methodologies presented in Table 7-1, advantages and limitations for each method are listed in Table 7-2.

Duration of sample running and sample treatment: As non-invasive detection methodology, IR spectroscopy is able to complete one sample scanning in a few seconds or minutes, without little or no sample preparation. Therefore, IR spectroscopy is more suitable for online or at line process control (Kamal and Karoui, 2015). In contrast, as laboratory methodology, LC separate sample compounds based on chemical characteristics after sample preparation, which would cost hours. For protein electrophoresis applied in Chapter 2, the whole procedure running cost 2-3 days, and there are preparation steps to extract protein before electrophoresis.

Accuracy, stability, and detection limit: As physico-chemical methods, LC methods are often considered as reference ones. HPLC often showed excellent performance on adulterant quantification in milk, as reported by (Krusa et al., 2000). In Chapter 3, low median CV (0.68-1.39%) of remarkable peptides intensities showed good reproducibility of this method. And low amounts (0.5%) of soy protein and HWP in total protein would be detected. 2-DE had a large CV variation (3.2-55.7%) of optical intensities for protein spots. And this method is only available on detection above 4% plant protein in total protein. Moreover, electrophoresis fingerprint is more qualitative than quantitative analysis technology, different from spectrum and chromatographic fingerprint (Zhang et al., 2011). In Chapter 6, IR methods failed to recognize adulteration below 6.25%. Spectrum variability between samples caused by milk composition difference may cover the real spectral difference induced by adulterants. In classification model, samples with HWP and HRP were classified correctly at dilution percentage above 6.25% in validation on sampling date, which means the calculated detection limit for ATR-MIR in Chapter 6 was about 8% of hydrolysed plant protein in total protein.

Data processing: This step is simple for electrophoresis. In Chapter 2, major differences of protein spots on gel were visible by naked eye. Recognition of some faint points may need help from special software. For chromatographic technology, matching of profile peaks, calculation of areas, data extraction, and further statistical analysis must be completed by expert on professional software with computers, like MaxQuant and Unscrambler used in Chapter 3. Spectral data analysis is complex for spectroscopy technology. Model validation is involved in chemometric approach, and these really rely on the chosen spectral technique and analysis methods, as well as proficient personnel (Lohumi et al., 2015). As mentioned in Chapter 5 and 6, processed spectral data by scatter correction or derivation, combined with PLS discrimination analysis and regression, classification model on adulterants types and quantification model on adulterants levels were developed respectively.

Cost and labour-consuming: As shown in Table 7-1, the most expensive instrument must be liquid chromatograph, then optical spectrometer and electrophoresis apparatus. And regents used LC and electrophoresis are also not cheap. In the view of labor-consuming, several manual steps involved in electrophoresis are not so friendly to technicians, especially for continuous measurements. For chromatogram, sample preparation before boarding need manual operation. With little or without sample treatments before measurements, procedure of IR is the simplest in the discussed three technologies.

2. Economic gain of adulterated samples in Chapter 6

To ensure a high quality of milk, the contents of protein and fat are taken into account by the dairy sector to fix the milk price, as showed in the formula (1; this formula is based on milk price fixed for a Belgian farm on October 2018). Therefore,

to increase their profitability, the temptation is high for some dairy farmers to elevate milk protein content by adulteration as much as possible.

$$\text{Price} = (\text{Fat}\% \times 318.62 + \text{Protein}\% \times 591.72)/100 \quad \ldots\ldots\ldots (1)$$

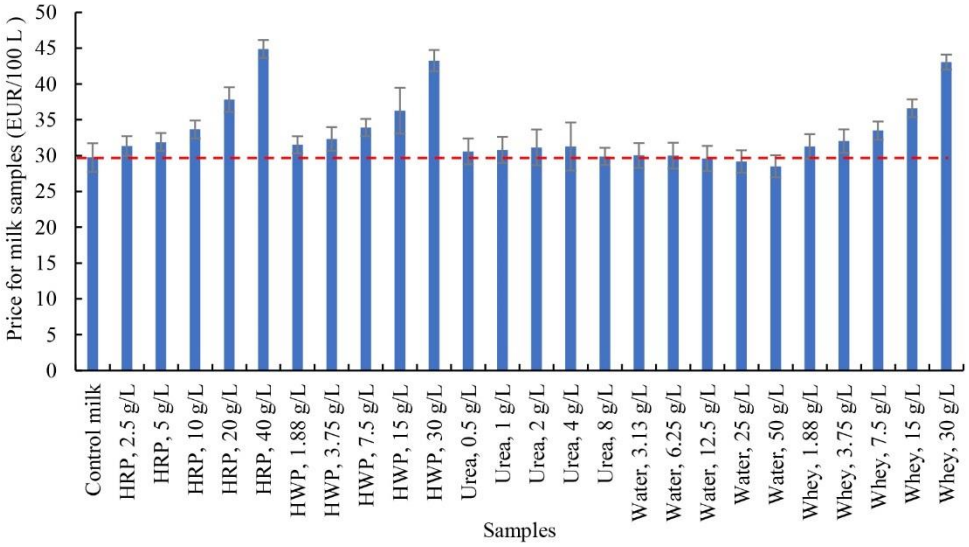


Figure 7-1: Price of samples in Chapter 6. Red dotted line indicated the price of control milk

For example, based on the contents of fat and protein predicted by NIR for the adulterated samples used in Chapter 6, unit prices for adulterated milk are presented in Figure 7-1. Compared with control milks (29.73 euro/100 L), higher prices were observed on samples adulterated with HRP, HWP, and whey, ranged 31.35-44.88, 31.52-43.26, and 31.27-43.06 euro/100 L, while similar prices for samples with urea (30.59-28.89 euro/100 L) or water (30.02-28.51 euro/100 L). As showed previously in Table 6-1, the addition of urea in milk did not raise the protein contents predicted by MPA, and protein contents for sample adulterated with water decreased. Therefore, there is no profit to spike the milk with urea and water based on the tested levels. In contrast, the protein contents of samples with HRP, HWP, and whey increased as adulterants level increased. As a result, prices of samples (similar fat contents) with HRP, HWP, and whey also increased with adulterants levels. Therefore, it is possible for samples with HRP, HWP, and whey to gain profit, of course, with low adulterants cost, whereas, addition of urea and water were proved to be useless to raise MPA prediction in this study. Nevertheless, developed ATR-MIR in Chapter 6 could identify samples with HRP above 5 g/L and HWP above 3.75 g/L, and NIR could recognize samples with whey above 7.5 g/L.

3. Combined methodologies for adulteration control

As mentioned above, LC and IR are often used to detect food adulteration. LC is powerful and robust, but not so favourable facing fast work flow and large number of samples. In contrast, rapid IR test is more and more popular on massive products monitoring, especially with the development of handheld and potable equipment. Combined with chemometrics, spectroscopy technique is suitable for samples screening. However, as secondary analytical methods, spectroscopies require accurate chemical and physical analyses as reference methods (Lohumi et al., 2015). Therefore, solid chemical methods based on LC should back up spectroscopy technologies. Suspicious samples from spectroscopy test should be submitted to confirmatory test in laboratory using LC. In addition, with the database contained spectroscopy data and reference values, not only the prediction of unknow samples using spectroscopy is realized, but the common equations across adulteration database in different instruments is possible after standardization of these instruments, as reported by Grelet et al. (2015).

4. Conclusions and Perspectives

This thesis tested and compared some methodologies on the detection of foreign protein added in bovine milk, especially (hydrolysed) plant protein. As a common proteomic technology, 2-DE has been proved to be an effective method to identify foreign protein in fluid milk by separation, with a detection limit of 4% of plant protein in total protein. Unfortunately, many manual steps are involved in this methodology. SDS-PAGE failed to detect the hydrolysed plant protein in fluid milk (i.e., no apparent protein lines in gels), but it turned out that high speed centrifugation (20 000 g× 60 min) reduce the cover effect of milk protein on minor foreign protein. Combined with PCA, peptides fingerprint revealed by HPLC-MS/MS differentiated control milk from adulterated samples, with a detection limit ranged from 0.5 to 2% of plant protein in total protein. However, identification of separated peptides using MS disabled to detect rice protein in samples with HRP. Coupled with PLS-DA, ATR-MIR spectra of skimmed samples could separate samples according to adulterants types (HRP, HSP, HWP, and whey) in cross-validation, with specificity of 87.5 - 100% and sensitivity of 80-100%. And well performed prediction on adulterants levels were also realized in cross- validation ($R^2 = 0.95$, RMSE = 2.25 g/L of full cross-validation). Moreover, combined with non-protein nitrogen (NPN) content in normal milk and predicted estimated NPN difference of adulterated samples from control milk, is was possible to classify samples adulterated with HRP, HSP, or HWP $\geq$ 3.8 g/L as adulterated samples. When variability of raw milk composition and external validation were considered, the classification model (PLS-DA) was not so successful to separate control milk form samples adulterated with HWP, HRP, whey, urea, and water, however, ATR-MIR showed better discriminability on HRP and HWP above 6.25%, while NIR showed its special discriminability on whey above 12.5%. And both ATR-MIR and NIR recognized urea adulteration in milk above 12.5%. Succeeding PLS regression model suggested better ATR-MIR and NIR prediction on HRP, HWP, and

urea, as well as foreign adulterants (without water) level in samples (validation R^2 of 0.87-0.99), and NIR presented better prediction on whey level in samples (validation $R^2 = 0.97$).

Based on the results in this thesis, the following points are recommended to improve the detection of foreign protein in milk adulteration:

1) As the potential methods recommended by official department, simplicity of procedures for LC or molecular method should be considered, for time- and labor- consuming, and too many manual steps involved.
2) Detection on hydrolyzed plant protein in milk should be developed. High hydrolysis degree of adulterants disturbs the identification using MS. A reliable method targeted on the characteristic and compounds of hydrolyzed plant protein may improve its detection.
3) As the most prospective methods, the effect of sample matrix on IR should be explored, although this method needs little sample preparation. In addition, simple and rapid removal of untargeted compounds before spectra acquisition may improve the efficiency of spectroscopy analysis.

5. References

Aernouts, B., E. Polshin, W. Saeys, and J. Lammertyn. 2011. Mid-infrared spectrometry of milk for dairy metabolomics: a comparison of two sampling techniques and effect of homogenization. Anal Chim Acta 705(1-2):88-97.

Cordewener, J. H. G., D. M. A. M. Luykx, R. Frankhuizen, M. G. E. G. Bremer, H. Hooijerink, and A. H. P. America. 2009. Untargeted LC-Q-TOF mass spectrometry method for the detection of adulterations in skimmed-milk powder. J Sep Sci 32(8):1216-1223.

Danezis, G. P., A. S. Tsagkaris, F. Camin, V. Brusic, and C. A. Georgiou. 2016. Food authentication: Techniques, trends & emerging approaches. TrAC Trends in Analytical Chemistry 85:123-132.

Du, L., W. Lu, Z. Cai, L. Bao, C. Hartmann, B. Gao, and L. Yu. 2018. Rapid detection of milk adulteration using intact protein flow injection mass spectrometric fingerprints combined with chemometrics. Food Chem 240(Supplement C):573-578.

Ellis, D. I., V. L. Brewster, W. B. Dunn, J. W. Allwood, A. P. Golovanov, and R. Goodacre. 2012. Fingerprinting food: current technologies for the detection of food adulteration and contamination. Chem Soc Rev 41(17):5706-5727.

Esslinger, S., J. Riedl, and C. Fauhl-Hassek. 2014. Potential and limitations of non-targeted fingerprinting for authentication of food in official control. Food Res Int 60(Supplement C):189-204.

Everstine, K., J. Spink, and S. Kennedy. 2013. Economically motivated adulteration (EMA) of food: common characteristics of EMA incidents. Journal of Food Protection 76(4):723-735.

Grelet, C., J. A. Fernández Pierna, P. Dardenne, V. Baeten, and F. Dehareng. 2015. Standardization of milk mid-infrared spectra from a European dairy network. Journal

of Dairy Science 98(4):2150-2160.
Issaq, H. J. and T. D. Veenstra. 2008. Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE): advances and perspectives. Biotechniques 44(5):697-+.
Jablonski, J. E., J. C. Moore, and J. M. Harnly. 2014. Nontargeted detection of adulteration of skim milk powder with foreign proteins using UHPLC-UV. J Agric Food Chem 62(22):5198-5206.
Kamal, M. and R. Karoui. 2015. Analytical methods coupled with chemometric tools for determining the authenticity and detecting the adulteration of dairy products: A review. Trends Food Sci Tech 46(1):27-48.
Karoui, R., A. M. Mouazen, E. Dufour, L. Pillonel, E. Schaller, D. Picque, J. De Baerdemaeker, and J.-O. Bosset. 2006. A comparison and joint use of NIR and MIR spectroscopic methods for the determination of some parameters in European Emmental cheese. Eur Food Res Technol 223(1):44-50.
Krusa, M., M. Torre, and M. L. Marina. 2000. A reversed-phase high-performance liquid chromatographic method for the determination of soya bean proteins in bovine milks. Anal Chem 72(8):1814-1818.
Lohumi, S., S. Lee, H. Lee, and B.-K. Cho. 2015. A review of vibrational spectroscopic techniques for the detection of food authenticity and adulteration. Trends Food Sci Tech 46(1):85-98.
Lopez-Tapia, J., M. R. Garcia-Risco, M. A. Manso, and R. Lopez-Fandino. 1999. Detection of the presence of soya protein in milk powder by sodium dodecyl sulfate capillary electrophoresis. J Chromatogr A 836(1):153-160.
Lu, W., J. Liu, B. Gao, X. Lv, and L. L. Yu. 2017. Technical note: Nontargeted detection of adulterated plant proteins in raw milk by UPLC-quadrupole time-of-flight mass spectrometric proteomics combined with chemometrics. J Dairy Sci 100(9):6980-6986.
Moore, J. C., J. W. DeVries, M. Lipp, J. C. Griffiths, and D. R. Abernethy. 2010. Total Protein Methods and Their Potential Utility to Reduce the Risk of Food Protein Adulteration. Compr Rev Food Sci F 9(4):330-357.
Rodriguez-Saona, L. E. and M. E. Allendorf. 2011. Use of FTIR for rapid authentication and detection of adulteration of food. Annu Rev Food Sci T 2:467-483.
Santos, P. M., E. R. Pereira, and L. E. Rodriguez-Saona. 2013. Application of hand-held and portable infrared spectrometers in bovine milk analysis. J Agr Food Chem 61(6):1205-1211.
Tessier, B., M. Schweizer, F. Fournier, X. Framboisier, I. Chevalot, R. Vanderesse, C. Harscoat, and I. Marc. 2005. Prediction of the amino acid composition of small peptides contained in a plant protein hydrolysate by LC–MS and CE–MS. Food Res Int 38(5):577-584.
Yang, Y. X., N. Zheng, J. H. Yang, D. P. Bu, J. Q. Wang, L. Ma, and P. Sun. 2014. Animal species milk identification by comparison of two-dimensional gel map profile and mass spectrometry approach. Int Dairy J 35(1):15-20.
Zhang, J., X. D. Zhang, L. Dediu, and C. Victor. 2011. Review of the current

application of fingerprinting allowing detection of food adulteration and fraud in China. Food Control 22(8):1126-1135.

8

Appendix: Research advances in milk production and detection by infrared spectroscopy (full text)

In this chapter, full text of "Research advances in milk production and detection by infrared spectroscopy" is presented. The infrared spectroscopy methodology is extensively applied to each link in the chain of milk production, from genetic selection of dairy cows to the assessment of milk quality. Many milk composition contents have been predicted by infrared spectroscopy, including fat, protein, lactose, and mineral elements. Infrared spectroscopy was also used to evaluate milk coagulation properties. Detection of adulterants in milk using spectroscopy methods has attracted great attention from many researchers. Genetic evaluation of spectral data was reported by recent studies, and a dietary effect on milk spectra was also observed. Moreover, milk spectra were found to be related to the energy status of dairy cows.

From Yang, J., N. Zheng, Y. Yang, Y. Zhang, and S. Li. 2016. Research advances in milk production and detection by infrared spectroscopy. Transactions of the Chinese Society of Agricultural Engineering, 32(17): 1-11 (*in Chinese*).

Abstract

Infrared spectroscopy (IR) can be used to determine the components in complexes. Instruments based on the Fourier transform have improved the accuracy and speed of IR analysis. This methodology is extensively applied to each aspect of milk production. Fat and protein contents in milk vary for different dairy farms, and many factors affecting milk quality contribute to the final acquisition price of raw milk. The determination of milk composition using IR provides a quick and comprehensive evaluation of milk quality. Unknown and undeclared adulterants in milk are a serious threat to consumers' health. Qualitative and quantitative analysis models provide a convenient method for detecting milk adulteration, resulting from spectrum variations due to these adulterants. Milk traits related to cows' health and robustness are very important for dairy farm management. Diagnosis of ketoacidosis and body energy status using IR instruments is helpful for selective breeding in dairy farms. This paper reviews the recent literature to evaluate the general trends in infrared spectroscopy applications for milk production. On the basis of introducing data processing and model building, this paper presents a review of the overseas and domestic literature on the evaluation of milk composition and milk coagulation properties using IR, especially for milk protein fractions and fatty acid composition. We compared the model performance for optical spectroscopy from different reports. The effect of the reference methods, sample size, and the units of model parameters were discussed in particular. Moreover, IR methods were found to be efficient for phenotype assessment and genetic selection based on these models. The variances in absorption on IR caused by adulterants not only indicated the appearance of milk adulteration, but also revealed the difference between cow milk and soy milk. Milk spectra were proven to be heritable in specified wavelengths, while other bands varied with different environmental factors. Many reports confirmed the correlation between cows' feed and milk optical characteristics. Although non-negligible random error and data variability existed in sampling, IR reflected the energy status of dairy cows with moderate accuracy. Mid-IR has been also studied as a potential tool to predict several milk traits related to cow health concerns, such as ketone bodies, which were closely related to cow fertility and milk production. IR was also used to predict methane emissions from cow digestive tracts. Emphasising the advantages of infrared spectroscopy analysis, we list potential challenges that exist in instrument setting, data collection, and model building. The objective of this paper is to highlight the application of infrared spectroscopy for milk traits, related to milk composition, quality, and dairy farm management. Considering the overall trends, we propose some future directions for research using this methodology in milk production, including prediction of trace nutrients, uniformity of reference methods and units, possibility of spectrum assessment, and diagnosis of disorders and fertility. With the future developments in these areas infrared methods would be more popular for milk composition determination, quality control, and dairy farm management, with high accuracy, efficiency, and convenience.

Key words: infrared spectroscopy, milk traits, adulteration, dairy farm management

1. Introduction

Compared with the time-consuming and labour-intensive laboratory reference methods, spectroscopy methods are becoming more popular for their rapid non-destructive testing. Infrared spectra (IR) are produced by the absorption change of infrared radiation interacting with molecular groups. According to the wavelength region, the spectra can be classified into near-infrared (NIR, 14 000-4 000 cm^{-1}), mid-infrared (MIR, 4 000-400 cm^{-1}), and far-infrared (FIR, 400-50 cm^{-1}). NIR and MIR are often used to detect, characterise, and quantify chemical components. NIR is the result of overtone and combination bands associated with the fundamental vibrations of hydrogen-containing functional groups. The band signal is relatively weak and suitable for direct analysis of highly-absorbed or strongly-scattered samples without pretreatment (Arbuckle et al., 1996; Rodriguez-Saona and Allendorf, 2011). MIR is the absorption band that is caused by the fundamental vibration of specific functional groups and can be used to identify the structure of organic components. The fingerprint area contains various structural information such as fats and proteins, and the ratio of band intensity to functional group concentration can be used for quantitative analysis (Paré and Bélanger, 1997; Rodriguez-Saona and Allendorf, 2011). Fourier transform (FT) devices improve the analysis speed and accuracy of spectroscopic techniques by resolving overlapping spectral bands, reducing bandwidth, and increasing peak height (Markovich and Pidgeon, 1991). Attenuated total reflectance (ATR) technology improves the accuracy of FTIR data, because multiple reflectance in samples increases the spectral response when compared with single-reflection crystals (Rodriguez-Saona and Allendorf, 2011).

NIR is widely used for the quantification of components in liquid milk and milk powder (Wu et al., 2008; Aernouts et al., 2011a; Inácio et al., 2011; Huang et al., 2014), identification of adulterants (Borin et al., 2006; Balabin and Smirnov, 2011; Huang et al., 2015), quality inspection (Al-Qadiri et al., 2008; Kong et al., 2013; Yazdanpanah and Langrish, 2013), and can be used for real-time on-line monitoring of raw milk production (Lyndgaard et al., 2012; Melfsen et al., 2012a; Santos et al., 2013b). MIR can not only accurately determine the milk composition, but also predict the milk fatty acids (FAs), protein components, and milk agglutination properties (De Marchi et al., 2014), and can be used to evaluate the genetic parameters of production traits (Bastin et al., 2012; Dagnachew et al., 2013b; Leclercq et al., 2013; Gustavsson et al., 2014). Negative energy balance, reproductive disorders, and ketosis seriously affect the performance of dairy cows and are the key to farm management. IR combined with feed intake, fatty acid (FA) composition and body ketone levels can provide reference information for the cow's body status and improve the efficiency of dairy management (Bastin et al., 2012; McParland et al., 2015). Through algorithm optimisation, spectral data standardisation between different instruments could be achieved. It has been proposed that a large spectral database based on networks across regions could be established to improve farm management (Grelet et al., 2015). This article focuses on the literature on IR applications for dairy production traits, quality inspection, and farm management in recent years, and provides some directions for future studies using IR.

2. Data processing and model establishment

Spectra are affected by many factors, such as the complexity and specificity of the absorption spectrum of chemical bonds, sample particle scattering and molecular interactions, variance in environmental conditions, differences in equipment performance, and so on. Band selection and data preprocessing are required to reduce the differences in data collection and improve model reliability (Zou et al., 2010; De Marchi et al., 2014). Common band selection methods include manual selection, multiple linear regression (MLR), successive projection algorithm (SPA), uninformative variable elimination (UVE), artificial neural networks (ANN), and genetic algorithms (GA). Data preprocessing methods mainly include scattering correction and its derivatives (Zou et al., 2010).

Cross-validation for the whole data set would overestimate the predictive power of an IR model, therefore a small additional test set is necessary for external validation. The number of samples in the calibration set should account for 50% or 75% of the total data (Bittante et al., 2014). Qualitative models classify samples according to absorption peaks based on pattern recognition methods such as correlation, distance, and discriminant analysis (Roggo et al., 2007). Commonly used evaluation parameters include the false positive rate, false negative rate, sensitivity, specificity, etc. (de Roos et al., 2007; Botelho et al., 2015). The quantitative model is based on the regression model derived from the relationship between spectral data and dependent variables in the calibration set and predicts the dependent variable using spectral data in the validation set. The root-mean-square error of prediction (RMSEP), calculated from the predicted value and the measured value, or the standard error of prediction (SEP) and determination coefficient (R^2) are used to evaluate the model performance (Zou et al., 2010). The ratio-performance deviation (RPD), range error ratio (RER), relative prediction error (RPE), and concordance correlation coefficient (CCC) are also important evaluation parameters (De Marchi et al., 2014). The qualitative analysis methods include the Mahalanobis distance, partial least squares discriminant analysis (PLS-DA), soft independent modelling of class analogy (SIMCA), and principal component analysis (PCA), etc., while the quantitative model commonly uses partial least square regression (PLSR), support vector machine (SVM), and ANN (Rodriguez-Saona and Allendorf, 2011; Domingo et al., 2014).

3. Milk composition detection

NIR predictions for milk composition are influenced by the spectral region, sample thickness, and measurement modes. The best accuracy for a NIR model was obtained with long wavelength bands (1 100 to 2 400 nm), 1 mm sample thickness, and the first derivative data transformation. For short wavelengths from 700 to 1 100 nm, the best accuracy for fat was obtained with a 10 mm sample, and for total protein with a 1 mm sample thickness. Lactose prediction was less affected by the sample thickness and spectral region (Tsenkova et al., 1999). NIR in reflectance mode resulted in accurate prediction of fat and crude protein in milk ($R^2 > 0.95$) and poor lactose prediction ($R^2 < 0.75$). In contrast, the transmittance spectra can achieve more accurate

predictions for these items, and the corresponding R^2 of prediction were 0.99, 0.93, and 0.88 (Aernouts et al., 2011a). Moreover, some studies have shown that NIR (851-1 649 nm) in diffusion reflectance has a similar or better prediction for fat, protein, and lactose in milk, compared with transmittance or transflectance mode (Melfsen et al., 2012b). NIR in diffusion reflectance not only accurately predicted fat, protein, and lactose in milk (R^2 = 0.99, 0.98, and 0.92, respectively, SEP = 0.09, 0.05, and 0.06), but also achieved a good prediction for urea and somatic cell counts (logarithmic transformation), R^2 of prediction was 0.82 and 0.85, and SEP was 19.3 mg/L and 0.18, respectively (Melfsen et al., 2012a). The presence of too many somatic cells in milk affected the NIR prediction for milk composition, so sorting of raw milk by somatic cell counts is necessary before dataset partition (Tsenkova et al., 2001). The prediction parameters for the main components in milk using NIR models from different studies are shown in Table 8-1, and NIR in diffuse reflectance is widely used. The good prediction performance of the NIR model makes it possible to assess fresh milk quality in real time, provide farmers with milk composition information and dairy cows' physiological status, and thereby to improve the efficiency of milk production (Kawasaki et al., 2008).

Table 8-1: Model performance of near-infrared spectroscopy for major milk components

Spectral mode	Fat		Protein		Lactose		References
	R^2	RMSEP	R^2	RMSEP	R^2	RMSEP	
Diffuse reflectance	0.977	0.154	0.960	0.134	-	-	Wang et al. (2015)
Transmittance	0.998	0.001	0.998	0.001	-	-	Zhao et al. (2014)
Fourier transform	0.995	0.136	0.975	0.195	-	-	Zhang (2010)
Transflectance	0.903	0.225	0.959	0.048	0.902	0.044	Yang et al. (2013)
Diffuse reflectance	0.998	0.09	0.98	0.05	0.92	0.06	Melfsen et al. (2012a)
Diffuse reflectance	0.95	0.25	0.83	0.26	0.72	0.15	Kawasaki et al. (2008)
Diffuse reflectance (Transmittance for lactose)	0.997	0.047	0.959	0.099	0.883	0.115	Aernouts et al. (2011a)

Pretreatment of milk samples, such as the addition of preservatives and homogenisation, would influence the MIR prediction. The addition of 0.02% potassium dichromate had little influence on the results of the MIR detection, whereas bromo-n-propylene glycol (0.02%) preserved milk had higher protein readings (a positive bias of about 0.01%) than potassium dichromate preserved or unpreserved milks. During cold preservation, uncorrected MIR readings for milk increased with

the storage time, the growth rate was higher for raw milk than for pasteurised milk, and the stability of the instrument zero was lower for raw milk than for pasteurised milk (Barbano et al., 2010). The prediction results for ATR on milk composition are better than those of high-throughput transmission spectra. Homogenisation was crucial to obtain a good fat prediction, but had little effect on the prediction of other components (Aernouts et al., 2011b). IR can predict not only the milk composition, but these spectroscopies can also detect the protein fraction, FA composition, and other trace substances in milk.

3.1. Protein

Determination of milk proteins related to the characteristic absorption of amide I and II bands at 1 700 to 1 500 cm^{-1}, and phosphate groups bound to casein at 1 100 to 1 060 cm^{-1}, other milk components (fat and lactose), and protein particles would affect the prediction of the PLS model for milk proteins (Etzion et al., 2004). A suitable region selection algorithm, SIMPLe-to-use Interactive Self-modeling Mixture Analysis, combined with IR can quantitatively predict the secondary structure of the polypeptide chain, and the correlation coefficient of cross-validation between the predicted and measured values of the α-helix and β-sheet was 0.86-0.98 (Bogomolov and Hachey, 2007). MIR may not be ideal for predicting individual milk protein composition with high accuracy. The R^2 of cross-validation (R^2cv) of MIR prediction of casein (CN), αs1CN, αs2CN, βCN, κCN, and γCN (g/L milk) in milk were 0.77, 0.66, 0.49, 0.53, 0.63, and 0.60, respectively, while the R^2cv for whey protein, alpha lactalbumin, and beta lactoglobulin (g/L milk) were 0.61, 0.31, and 0.64, respectively (Bonfatti et al., 2011). Other studies have similar predictions of whey protein and its fractions (alpha lactalbumin, and beta lactoglobulin), and poor prediction of total and individual caseins, using raw MIR spectra (De Marchi et al., 2009a; Rutten et al., 2011a), whereas there are some reports that predict the total CN with validation $R^2 > 0.90$ (Luginbuhl, 2002). The difference between these models may be associated to the reference methods for protein determination used in these studies (De Marchi et al., 2014). The MIR prediction for milk protein composition can be used to estimate breeding values and improve protein composition on a genetic level (Rutten et al., 2011a). FTIR combined with PLS can distinguish the milk produced by goats with two weak haploids from others, thereby selection of goats with high casein expression or screening for milk samples with high casein content is possible (Berget et al., 2010). FTIR prediction of β-LG genotypes showed a repeatability of 0.85, and it can improve the percentage of correctly predicted β-LG genotypes, in combination with pedigree information and derived genotypes (Rutten et al., 2011b).

3.2. Fatty Acids Composition

There are two absorption bands for milk fat in the MIR region, fat A at 5.73 μm and fat B at 3.48 μm, which involved the stretch of C=O and C–H, respectively (Biggs and McKenna, 1989). Fat B MIR predictions increased and fat A MIR prediction decreased relative to reference chemistry with increasing FA chain length. When MIR fat prediction of fat B was corrected according to unsaturation variation between

samples, fat B had a positive correlation with the FA chain length (correlation coefficient was 0.42-0.89); when the corrected ratio of fat B for unsaturation was 45:55, fat A gave the best fit between MIR prediction and the reference (Kaylegian et al., 2009). Oleic acid and linoleic acid presented different spectra in the MIR region. Oleic acid has two characteristic peaks at 1 119 and 1 091 cm^{-1}, while the characteristic peaks of linoleic acid appear at 1 048, 1102, and 1 121 cm^{-1} (Yang, 2011). The IR prediction of milk FAs in different studies are shown in Tables 8-2, 8- 3, and 8-4. Among unsaturated FAs, a better prediction accuracy for c9C18:1 is observed; meanwhile, the prediction accuracy for saturated FAs and monounsaturated FAs is greater than that of polyunsaturated FAs, which may be in line with individual and grouped FA concentration in milk (Soyeurt et al., 2006; Rutten et al., 2009; De Marchi et al., 2011). There are two ways of expressing FAs in milk, namely the concentration of FAs in milk (g/L milk or g/kg milk) and the FA content in fat (g/kg total FAs). The accuracy of MIR predictions expressed as the FA concentration are better than those of models expressed as FA content (Soyeurt et al., 2006; De Marchi et al., 2014), which is similar to the NIR prediction for liquid milk (Coppa et al., 2014). However, the comparison of prediction accuracy between MIR and NIR for FAs in oven-dried milk varied with FAs and expression. When expressed as g/kg of milk, the accuracy of NIR prediction was worse than MIR for almost all FAs. When expressed as g/100 g total FAs, MIR and NIR shared a similar prediction accuracy for the group of even-chain saturated FA, odd-chain FA, unsaturated FA, conjugated linoleic acid, n-3 FA, and c9C18:1/C16 ratio; while monounsaturated FA, n-6/n-3 ratio, polyunsaturated FA (PUFA), and n-6 FA were better predicted by NIR (Coppa et al., 2014). High levels of FAs in milk, such as even-chain FAs, could achieve good MIR prediction fitting with measured values, no matter the expression of FAs (Soyeurt et al., 2006). Compared with predictions for liquid milk, NIR quantification of milk FA was more accurate or similar for oven-dried milk (Coppa et al., 2010; Coppa et al., 2014), but the reliability decreased for thawed liquid milk (Coppa et al., 2014). MIR prediction was also used to estimate heritability and correlation of FAs in goat milk (Maroteau et al., 2014). Strong relationships between the sample size of calibration and validation R^2, as well as strong genetic correlations were observed. As the calibration number increased, the variation range of the validation R^2 and the genetic correlation coefficient gradually narrowed. When there were 1 000 samples in calibration, the genetic correlation changed within a range of 0.1 (Rutten et al., 2010).

Table 8-2: Model prediction of infrared spectroscopy for milk fatty acids (saturated fatty acid)

Sample size	Unit	C4:0	C6:0	C8:0	C10:0	C12:0	C14:0	C15:0	C16:0	C17:0	C18:0	References
267	g/L of cow milk	-	-	0.74 (0.07)	0.73 (0.19)	0.75 (0.25)	0.77 (0.6)	0.63 (0.07)	0.70 (1.59)	0.56 (0.03)	0.65 (0.75)	De Marchi et al. (2011) [1]
600	g/L of cow milk	0.51 (0.08)	0.52 (0.04)	0.59 (0.02)	0.64 (0.04)	0.74 (0.02)	0.82 (0.05)	0.40 (0.01)	0.82 (0.17)	-	0.69 (0.13)	Soyeurt et al. (2006) [2]
600	g/kg of milk fat	0.39 (1.60)	0.41 (0.98)	0.46 (0.50)	0.53 (0.90)	0.64 (0.53)	0.67 (1.14)	0.53 (0.2)	0.50 (3.5)	-	0.09 (2.77)	
517	g/L of cow milk	0.94 (0.01)	0.97 (0.00)	0.97 (0.00)	0.96 (0.01)	0.96 (0.01)	0.97 (0.02)	-	0.95 (0.08)	0.89 (0.00)	0.90 (0.05)	Soyeurt et al. (2011) [2]
3 660	g/L of cow milk	0.91 (0.10)	0.96 (0.20)	0.94 (0.50)	0.92 (0.10)	0.85 (0.30)	0.94 (0.3)	-	0.94 (0.10)	-	0.82 (0.70)	Rutten et al. (2009) [3]
3 660	g/kg of milk fat	0.55 (0.00)	0.73 (0.30)	0.73 (0.60)	0.75 (0.20)	0.68 (0.30)	0.73 (0.3)	-	0.71 (0.00)	-	0.51(1.20)	
238-241[a]	g/L of cow milk	0.93 (0.006)	0.96 (0.003)	0.96 (0.002)	0.95 (0.007)	0.95 (0.008)	0.94 (0.024)	-	0.94 (0.066)	-	0.84 (0.041)	
98-104[b]	g/L of cow milk	0.61 (0.01)	0.86 (0.004)	0.89 (0.003)	0.85 (0.011)	0.82 (0.018)	0.84 (0.03)	-	0.82 (0.111)	-	0.49 (0.054)	Ferrand-Calmels et al. (2014) [4]
135-140	g/L of sheep milk	0.93 (0.01)	0.97 (0.005)	0.96 (0.008)	0.93 (0.041)	0.97 (0.019)	0.96 (0.045)	-	0.94 (0.091)	-	0.83 (0.061)	
215-229	g/L of goat milk	0.96 (0.004)	0.95 (0.004)	0.97 (0.004)	0.98 (0.013)	0.92 (0.013)	0.93 (0.023)	-	0.96 (0.042)	-	0.86 (0.034)	

Table 8-2 continued

Sample size	Unit	C4:0	C6:0	C8:0	C10:0	C12:0	C14:0	C15:0	C16:0	C17:0	C18:0	References
154	g/L of cow milk	0.87 (0.009)	0.97 (0.003)	0.97 (0.002)	0.95 (0.008)	0.95 (0.011)	0.83 (0.040)	0,67 (0.005)	0.91 (0.087)	0.74 (0.002)	0.75 (0.048)	Ferrand et al. (2011) [2]
1 167-1 187	g/L of cow milk	0.93 (0.008)	0.96 (0.005)	0.96 (0.003)	0.96 (0.008)	0.95 (0.01)	0.95 (0.028)	-	0.97 (0.068)	0.89 (0.003)	0.90 (0.045)	Maurice-Van Eijndhoven et al. (2013) [2]
279-344	g/kg of total FA	0.66 (0.42)	0.88 (0.21)	0.90 (0.13)	0.91 (0.34)	0.89 (0.41)	0.88 (1.07)	0.53 (0.14)	0.91 (2.20)	0.65 (0.08)	0.80 (1.31)	Coppa et al. (2010) [5]

Notes: Parameters of models in this table are as follows: 1) correlation coefficient of cross-validation (RMSE), 2) determination coefficient of cross-validation (SE), 3) determination coefficient of validation (bias/mean × 100), 4) determination coefficient of validation (residual error), 5) determination coefficient of validation (prediction error); 1, 2, 3, 4 used the mid-infrared spectroscopy model; 5 used the near-infrared spectroscopy model, spectra scanned for oven dried milk

a) data from MilkoScan FT6000, b) data from Bentley FTS

Table 8-3: Model prediction of infrared spectroscopy for milk fatty acids (unsaturated fatty acids and fatty acid groups)

Sample size	Unit	c9C14:1	c9C16:1	c9C18:1	c11C18:1	c9c12C18:2	C18:3n-3	c9t11C18:2	References
267	g/L of cow milk	0.68 (0.08)	0.60 (0.11)	0.73 (1.13)	0.59 (0.04)	-	0.51 (0.04)	0.58 (0.04)	De Marchi et al. (2011)[1]
600	g/L of cow milk	0.07 (0.01)	0.65 (0.02)	-	-	0.62 (0.02)	0.14 (0.01)	0.07 (0.02)	(Soyeurt et al., 2006) [2]
600	g/kg of milk fat	0.23 (0.28)	0.37 (0.37)	-	-	0.11 (0.44)	0.20 (0.20)	0.34 (0.37)	
517	g/L of cow milk	0.68 (0.01)	0.71 (0.01)	0.97 (0.05)	-	0.74 (0.01)	0.71 (0.01)	0.74 (0.01)	Soyeurt et al. (2011) [2]
3 660	g/L of cow milk	-	-	0.92 (0.30)	0.27 (0.10)	0.36 (0.90)	0.45 (3.30)	0.58 (1.00)	Rutten et al. (2009) [3]
3 660	g/kg of milk fat	-	-	0.84 (0.50)	0.22 (0.40)	0.28 (0.60)	0.38 (2.80)	0.56 (1.10)	
238-241[a]	g/L of cow milk	-	-	0.96 (0.039)	-	0.77 (0.006)	0.85 (0.004)	0.82 (0.003)	Ferrand-Calmels et al. (2014) [4]
98-104[b]	g/L of cow milk	-	-	0.86 (0.063)	-	0.75 (0.006)	0.81 (0.003)	0.64 (0.003)	
135-140	g/L of cow milk	-	-	0.97 (0.057)	-	0.49 (0.012)	0.74 (0.007)	0.91 (0.011)	
215-229	g/L of cow milk	-	-	0.95 (0.037)	-	0.89 (0.007)	0.79 (0.003)	0.71 (0.003)	

Table 8-3 continued

Sample size	Unit	c9C14:1	c9C16:1	c9C18:1	c11C18:1	c9c12C18:2	C18:3n-3	c9t11C18:2	References
154	g/L of cow milk	-	-	-	-	0.76 (0.006)	0.85 (0.003)	0.66 (0.004)	Ferrand et al. (2011) [2]
1 167-1 187	g/L of cow milk	0.78 (0.007)	0.78 (0.011)	-	-	-	-	-	Maurice-Van Eijndhoven et al. (2013) [3]
279-344	g/kg of total FA	0.57 (0.22)	0.44 (0.25)	0.93 (1.77)	0.29 (0.13)	0.34 (0.28)	0.48 (0.16)	0.73 (0.87)	Coppa et al. (2010) [5]

Notes: Parameters of models in this table are as follows: 1) correlation coefficient of cross-validation (RMSE), 2) determination coefficient of cross-validation (SE), 3) determination coefficient of validation (bias/mean × 100), 4) determination coefficient of validation (residual error), 5) determination coefficient of validation (prediction error); 1, 2, 3, 4 used the mid-infrared spectroscopy model; 5 used the near-infrared spectroscopy model, spectra scanned for oven dried milk

a) data from MilkoScan FT6000, b) data from Bentley FTS

Table 8-4: Model prediction of infrared spectroscopy for milk fatty acids (fatty acid groups)

Sample size	Unit	SCFA	MCFA	LCFA	SFA	MUFA	PUFA	UFA	References
267	g/L of cow milk	-	0.73 (2.66)	0.76 (1.94)	0.72 (2.97)	0.74 (1.39)	0.64 (0.22)	0.71 (1.57)	De Marchi et al. (2011)[1]
600	g/L of cow milk	-	-	-	0.94 (0.20)	0.85 (0.22)	0.39 (0.04)	0.66 (0.34)	Soyeurt et al. (2006) [2]
600	g/kg of milk fat	-	-	-	0.63 (3.75)	0.52 (4.10)	0.10 (0.74)	0.63 (3.75)	
517	g/L of cow milk	0.98 (0.02)	0.98 (0.09)	0.98 (0.09)	1.00 (0.05)	0.99 (0.04)	0.85 (0.02)	0.99 (0.04)	Soyeurt et al. (2011) [2]
3 660	g/L of cow milk	0.95 (0.00)	0.97 (0.00)	-	-	-	-	-	Rutten et al. (2009) [3]
3 660	g/kg of milk fat	0.82 (0.30)	0.77 (0.10)	-	-	-	-	-	
238-241[a]	g/L of cow milk	-	-	-	1.00 (0.035)	0.97 (0.037)	0.76 (0.01)	0.98 (0.038)	Ferrand-Calmels et al. (2014) [4]
98-104[b]	g/L of cow milk	-	-	-	0.96 (0.09)	0.89 (0.068)	0.60 (0.01)	0.83 (0.10)	
135-140	g/L of cow milk	-	-	-	1.00 (0.049)	0.99 (0.044)	0.96 (0.015)	0.99 (0.048)	
215-229	g/L of cow milk	-	-	-	0.99 (0.043)	0.96 (0.037)	0.92 (0.01)	0.97 (0.039)	

Table 8-4 continued

Sample size	Unit	SCFA	MCFA	LCFA	SFA	MUFA	PUFA	UFA	References
154	g/L of cow milk	-	-	-	0.99 (0.045)	0.97 (0.044)	0.62 (0.010)	-	Ferrand et al. (2011) [2]
1167-1 187	g/L of cow milk	0.96 (0.02)	0.98 (0.086)	-	1.00 (0.051)	-	-	-	Maurice-Van Eijndhoven et al. (2013) [3]
279-344	g/kg of total FA	-	-	-	0.97 (1.94)	0.97 (1.81)	0.85 (0.87)	0.97 (2.23)	Coppa et al. (2010) [5]

Notes: Parameters of models in this table are as follows: 1) correlation coefficient of cross-validation (RMSE), 2) determination coefficient of cross-validation (SE), 3) determination coefficient of validation (bias/mean×100), 4) determination coefficient of validation (residual error), 5) determination coefficient of validation (prediction error); 1, 2, 3, 4 used the mid-infrared spectroscopy model; 5 used the near-infrared spectroscopy model, spectra scanned for oven dried milk

a) data from milkoscan FT6000, b) data from Bentley FTS

SCFA: short chain fatty acids; MCFA: medium chain fatty acids; LCFA: long chain fatty acids; SFA: saturated fatty acid; MUFA: monounsaturated fatty acids; PUFA: polyunsaturated fatty acids; UFA: unsaturated fatty acids.

3.3. Other components

MIR can be used to predict cheese production. Favourable predictions were observed for the yield of total solid cheese and fresh cheese, with R^2cv of 0.95 and 0.83, respectively, and promising results were obtained for the recovered protein, total solids, and energy (R^2cv were 0.81, 0.86, and 0.76, respectively) (Ferragina et al., 2013). Visible and short wavelength NIR diffuse reflectance spectroscopy (600-1 000 nm) can be used to monitor spoilage of pasteurised skimmed milk by predicting the bacterial counts and pH of milk (R^2cv = 0.99 and 0.99, SEP = 0.34 cfu/mL and 0.031, respectively) (Al-Qadiri et al., 2008). FT-MIR can also accurately determine the titration acidity of milk, with R^2cv of 0.96, RMSE of the cross validation set of 0.72°T, and the RPD of 5.1 (Calamari et al., 2016). The pretreated NIR spectra combined multivariate regression model could predict the IgG level in colostrum, where the R^2 of the calibration set and cross validation set were 0.95 and 0.94, respectively (Rivero et al., 2012). The potential for MIR prediction of lactoferrin was also confirmed by studies, with R^2 of cross validation and external validation of 0.71 and 0.60 (Soyeurt et al., 2012), and there was a positive correlation between predicted lactoferrin and somatic cell scores, but a negative genetic correlation between predicted lactoferrin and milk yield was also observed (Soyeurt et al., 2007). When MIR combined with atomic absorption spectrometry was used to predict major mineral elements Ca, K, Mg, Na, and P in milk, only Ca, Na, and P showed the sufficient R^2cv (0.80, 0.70, and 0.79) for potential application. Finally, potential application of Ca and P equations were confirmed, whose R^2 of external validation were 0.97 and 0.88 (Soyeurt et al., 2009). Accurate MIR prediction for Ca and P was confirmed by other studies, and the contents of these elements were closely related to the agglutination traits of milk (Toffanin et al., 2015). FT-MIR can predict tetracycline levels in milk, validation R^2 reached 0.85-0.89, the detection range was 4-2 000 μg/kg, and SEP was from 89 to 387 μg/kg (Sivakesava and Irudayaraj, 2002).

4. Evaluation of cheese making performance

As the demand for and production of cheese continues to increase, prediction of milk's agglutination performance and corresponding processing index becomes crucial. The experimental NIR data could fit the model of the whole coagulation process and the models of three stages of milk agglutination: κCN hydrolysis, protein micelle polymerisation, and colloid formation, very well (R^2 > 0.99) (Lyndgaard et al., 2012). The genetic correlation between NIR prediction and measured rennet coagulation time (RCT) and coagulation block hardness (a_{30}) after 30 minutes of chymosin addition, was 0.97 and 0.92, respectively (Cecchinato et al., 2013), while the parameters of corresponding MIR prediction ranged from 0.91-0.96 and 0.71-0.87, respectively (Cecchinato et al., 2009). The R^2 and the standard error of cross validation (SECV) for MIR models of milk agglutination indicators in different studies are shown in Table 8-5. Different reference methods, i.e. computerised renneting meter (CRM) and Formagraph (FOR), had a limited impact on R^2cv of the RCT prediction model (0.63, 0.64, and 0.61- 0.76), and a medium impact on R^2 of the a_{30} prediction

model (0.36, 0.35, and 0.50- 0.70). Three regions of MIR, 1 600 to 900 cm^{-1}, 3 040 to 1 700 cm^{-1}, and 4 000 to 3 470 cm^{-1}, can be used to improve the predictions of milk coagulation properties, titratable acidity, and pH, respectively. An approximate prediction was achieved by MIR models for titration acidity (R^2 = 0.66), while predictions of RCT and pH can only differentiate high and low values (R^2 range 0.59-0.62) (De Marchi et al., 2009b). At the same time, there is no specific MIR information to distinguish coagulating milk from non-coagulating milk (De Marchi et al., 2013). In the external validation, the proportions of variance explained by prediction models for pH, RCT, and heat coagulation time were 71%, 55%, and 46%, respectively; all regression models could not be used to analyse traits (regression coefficients less than 1 and RPD less than 2), but the CCC in external validation ranged from 0.63 (heat coagulation time) to 0.84 (pH), which means the MIR predictions can be used as a screening tool (Visentin et al., 2015). Addition of bronopol to cow's milk can improve the prediction of the MIR model. R^2cv for RCT models were 0.73, 0.55, and 0.41, 0.29 after freezing for 4 d and 8 d with and without preservatives, respectively (Dal Zotto et al., 2008). When MIR predictions were used to evaluate genetic parameters of the milk coagulation index, it was found that the heritability of predicted RCT (0.26) and clot coagulation time (k_{20}) (0.31) analysis was close to measured milk coagulation properties in the literature, while the heritability of predicted a_{30} was higher than reported measured values (Chessa et al., 2014). The repeatability of the MIR prediction of RCT and a_{30} was 92.8% and 95.8%, which was close to that of the FOR-reference method (98.6 and 95.8%), but the reproducibility was only 67.3% and 71.9%, respectively (Penasa et al., 2015). In another study, CRM was used as the reference method, whose repeatability for MIR prediction of RCT and a_{30} under different preserving conditions was 95.7% and 77.3%, respectively, and the reproducibility was 93.5% and 64.6%, respectively (Dal Zotto et al., 2008). This shows that different reference methods also have great influence on the MIR prediction of milk coagulation properties.

Table 8-5: Milk coagulation properties found by infrared spectroscopy in different references

Sample size	Test methods	RCT		a_{30}		k_{20}		References
		R^2	SECV	R^2	SECV	R^2	SECV	
1049	CRM	0.63	2.36	0.36	6.86	-	-	(De Marchi et al., 2009b)
79	CRM	0.64	2.16	0.35	6.82	-	-	(Dal Zotto et al., 2008)
147-319	FOR	0.76	7.05	0.70	7.68	0.72	3.54	(De Marchi et al., 2013)
250	FOR	0.65	2.77	0.68	5.11	0.49	1.81	(Chessa et al., 2014)
378-450	FOR	0.61	5.64	0.50	11.32	0.59	0.39	(Visentin et al., 2015)

Notes: RCT: rennet coagulation time; a_{30}: curd firmness 30 min after rennet addition; k_{20}: curd firming time; CRM: computerised renneting meter (Polo Trade, Monselice, Italy); FOR:

Formagraph (Foss Electric A/S, Hillerød, Denmark); SECV: standard error of cross validation.

5. Milk quality inspection

Combined with SIMCA, short-wave NIR can distinguish milk stored for 30 hours at different temperatures (6, 21, and 37°C) from control samples, with an accuracy of about 90% (Al-Qadiri et al., 2008). FT-NIR combined with principal component analysis can accurately recognise different brands of milk with an accuracy of 100% (Jin et al., 2016). And coupled with Fisher's multi-class linear discriminant analysis, FT-NIR can identify milk adulteration with plant cream, vegetable protein, and starch, and the correct rate was achieved in more than 94% of cases (Li and Ding, 2010). Compared with 1 704-1 400 cm^{-1} in the MIR region, 4 800-4 200 cm^{-1} in NIR was more sensitive to urea adulterated milk; for unknown samples, the prediction R^2 reached 0.999, and RMSEP was 0.219 g/L (Yang et al., 2012). Based on changes in fat and protein contents in milk within 52 hours, NIR could indicate changes in milk quality (Shi, 2014). Based on the prediction of pH and acidity of goat milk, NIR can also be used to evaluate the freshness of goat milk (Chu, 2012). Since phenolic compounds were stable during fermentation and manufacturing processing, it was possible to discriminate probiotic milk samples according to the type of extract added and to evaluate the 'stability' of the product using NIR spectra combined with multivariate analysis (Aliakbarian et al., 2015). The adulterant levels in milk affected the discrimination of the NIR models, and the combination of non-linear pattern recognition and NIR could be useful for the identification and authentication of raw cow milks (Zhang et al., 2014). Nevertheless, some results indicated that the MIR system was superior to the NIR system in monitoring milk adulteration for additives such as water, whey protein, synthetic milk, synthetic urea, urea, and hydrogen peroxide (Santos et al., 2013b). A comprehensive index Q was constructed using milk indexes detected by a FTIR analyser, total solid-fat, ice-lactose, and lactose parameters were mainly included, and the addition of butter (> 0.058 g/100 g), gelatin hydrolysate (> 0.020 g/100 g), ammonium chloride (> 0.395 g/100 g), melamine (> 0.310 g/100 g), urea (> 0.443 g/100 g), sucrose and maltodextrin (> 0.024 g/100 g), whey (> 0.072 g/100 g), and milk powder and water (> 0.500 g/100 g) to milk could be recognised (Liu et al., 2015). FTIR could quantitatively detect the spiked level of baking soda, sodium citrate, and lactalbumin in milk, with all calibration R^2 above 0.91 and the detection limits of 0.015%, 0.017%, and 3.9%, respectively (Cassoli et al., 2011). ATR-MIR combined with PLS-DA was able to detect the presence of water, starch, sodium citrate, formaldehyde, and sucrose in milk, and the detection range was 0.5% to 10.0% (w/v) (Botelho et al., 2015). ATR-MIR also detected whey protein, hydrogen peroxide, synthetic urine, urea, and synthetic milk adulterated milk, and SEP was 2.33, 0.06, 0.41, 0.30, and 0.014 g/L, respectively, and the detection limits were 7.5, 0.019, 0.78, 0.78, and 0.1 g/L respectively (Santos et al., 2013a). Combined with PLS, single-beam ATR-FTIR can quickly predict melamine content in milk with the limits of detection and quantitation 2.5 and 15 mg/kg, respectively (Jawaid et al., 2013). If the correct data processing and multivariate algorithm was applied in the developed model, the detection limit of melamine for IR prediction could be less than

1 mg/kg, and a non-linear relationship was found between melamine content and IR response (Balabin and Smirnov, 2011; Domingo et al., 2014).

MIR was also sensitive to adulteration between different milks. The ratio of cow's milk, goat's milk, and sheep's milk in their mixture could be quantitatively predicted by FTIR. The prediction R^2 for a binary mixture was 0.91-0.98, RMSEP was 3.95-8.03%; the prediction R^2 for a ternary mixture was 0.92-0.97, RMSEP ranged from 3.36% to 6.40% (Nicolaou et al., 2010). A main peak located at 1745 cm^{-1}, related to the degree of sugar carboxyl methyl esterification, was observed on the FTIR comparison between goat's and sheep's milk, and corresponding hierarchical and discriminant analyses showed goat samples could be separated from sheep samples (Pappas et al., 2008). Differences between soy milk and cow-buffalo milk, as well as their mixture at different ratios were centred on the MIR region at 1 680 to 1 058 cm^{-1}. PCA indicated that the addition above 5% of soybean milk to milk showed a significant difference from control milk. Based on the absorption of 1 472 to 1 241 cm^{-1}, the multivariate linear regression analysis showed that validation R^2 was 0.92 and the SEP was 7.56 for soybean milk levels in the mixture (Jaiswal et al., 2015).

6. Milk Spectral Genetic Characteristics and Dietary Effects

The most individual waves in the 5 000-930 cm^{-1} region of FTIR transmittance spectrum for bovine milk was heritable, and transmittance of contiguous FTIR waves correlated with phenotypic variation more than with genetic variation; transmittances on short-wavelength infrared (SWIR)- mid-wavelength infrared (MWIR) (3 669-3 052 cm^{-1}) and MWIR2 (1 698-1 586 cm^{-1}), which was related to the absorption of water in this area, were characterised by very high phenotypic and genetic variability in the transmittance within each wave, as well as low heritability estimates, while SWIR (5 000-3 673 cm^{-1}), MWIR1 (3 048 - 1 701 cm^{-1}) and MWIR- long-wavelength infrared (LWIR, 1 582-930 cm^{-1}) showed low phenotypic and genetic variability for individual waves, and heritability estimates of transmittance on the SWIR (5000-3 673 cm^{-1}) region presented a rare cyclic pattern (Bittante and Cecchinato, 2013). The reduced dimension analysis of 1 060 data points in the milk spectrum showed that the spectrum can be expressed as 46 traits, among these 8 traits had a heritability larger than 0.1, 25 traits showed a permanent environmental variance greater than genetic variance, and 3 infrared regions exhibited moderate to high heritability (Soyeurt et al., 2010). The heritability of goat milk spectral variables varied from 0.018 to 0.408, the permanent environmental effect variance was 0.002 to 0.184 for phenotypic spectral variation, and the spectral regions associated with the milk components (fat, lactose, and protein), 1 300-1 030, 1 600-1 500, 1 800-1 700, and 3 000-2 800 cm^{-1}, showed high to moderate heritability; some spectral regions were greatly affected by herd test-day variation, which could be helpful for herd management (Dagnachew et al., 2013a). When compared with indirect measurement (the spectral prediction of milk components combined with the pedigree information to calculate the estimated breeding value), the direct measurement (genetic analyses of spectral variables) reduced the prediction error variance of fat, lactose, and protein

by 3.73, 4.07, and 7.04% respectively, and corresponding genetic gains were 2.99, 2.98, and 4.85%, respectively. Therefore, the estimated breeding value of milk FTIR spectra was more accurate than the single-trait animal model analyses (indirect measurement) on phenotypes predicted separately from the spectra (Dagnachew et al., 2013b).

Compared with other factors, the difference between milk spectra is more affected by dairy cows' diets. The type of diet had little effect on NIR prediction of milk fat, but the prediction accuracy of protein was significantly affected (Purnomoadi et al., 1999). NIR could discriminate no-pasture samples from pasture milk, even when pasture occupied only 30% of the diet (5.4% cross validation error), and the stabilised error (2.5% error) was observed when pasture exceeded 70%, however this was not enough to reliably trace the geographic origin of milk production (Coppa et al., 2012). NIR combined with PLSR could separate ewe milk according to different feeding systems (grazing, box feeding) (Mouazen et al., 2009). A good discrimination between milk from ewes fed soybean meal and scotch bean meal could be obtained based on spectral differences in the 3 000-2 800 cm^{-1} and 1 500-900 cm^{-1} regions, but the spectra cannot distinguish milk from different lactation stages (Maâmouri et al., 2008). In the external validation of the FTIR combined with the PLS-DA model, the sensitivity and specificity of classification of milk from cows with or without fresh pasture in their diet were 88% and 83%, respectively, and the classification accuracy for the organic and conventional samples was 80% and 94% respectively. Milk samples from grazing cows and barn-fed cows can only be accurately distinguished in the whole sample set (Capuano et al., 2014). MIR can distinguish milk from a hay-pasture based system from those from a corn silage-forage based system, but it is not possible to distinguish milk between cow breeds and between grazing altitudes (Valenti et al., 2013).

7. Energy intake, health status, and methane emission diagnosis in dairy cows

Available MIR prediction for the energy status of dairy cows was realised in recent studies. The accuracy of MIR prediction for direct energy balance, body energy content, and the energy intake of dairy cows ranged from 0.47-0.69, 0.51-0.56, and 0.76-0.80, respectively, but the large random error in the calculation using gap filled data directly affected the accuracy of MIR prediction (McParland et al., 2012). The accuracy of cross validation of MIR predictions for daily changes in body condition scores and body weight for cows were 0.77 and 0.70, respectively, and the heritability of two predictions were similar to the estimated heritability (0.07, 0.06, and 0.07, 0.08), which means that IR could be helpful for monitoring the energy intake and energy balance of dairy cows (McParland et al., 2015). The correlation between the measured residual feed intake and measured energy balance across lactation was 0.85, and the correlation between corresponding MIR predictions was 0.65. Therefore, when combined with the developed prediction equations, milk MIR spectra data collected from individual cows throughout lactation could predict energy intake and feed efficiency in dairy cows (McParland et al., 2014).

MIR is also an alternative tool to screen ketosis in cows. When the acetone in samples ranged from 0 to 2.8 mM, the obtained prediction R^2 and RMSE were 0.81 and 0.27 mM, respectively (Hansen, 1999). The characteristic absorption of acetone is located at 1 450-1 200 cm^{-1}. When the threshold value of subclinical ketosis was 0.4-1.0 mM, the sensitivity and specificity of the classification model (FTIR coupled with PLS) was 95-100% and 96-100%, respectively. When the assumed prevalence ranged from 10% to 30%, the positive predictive value and negative predictive value were ≥ 76% and ≥ 98%, respectively (Heuer et al., 2001). The correlation coefficients between MIR prediction and measured values of β-hydroxybutyric acid and acetone in milk were 0.85 and 0.79, respectively. If 0.15 mM acetone and 0.10 mM β-hydroxybutyric acid in milk were set as the threshold for subclinical ketosis, high concentrations of acetone or beta-hydroxybutyric acid were detected with a sensitivity of 69-70%, a specificity of 95%, false positives of 25-27%, and false negatives of 6-7% (de Roos et al., 2007). Moreover, FTIR prediction of β-hydroxybutyric acid and acetone could diagnose cows with hyperketonemia in early lactation with higher accuracy than fat/protein ratio, but the high proportion of false-positive tests for these indicators should be improved in future studies (van Knegsel et al., 2010).

The measured CH_4 emission at day 0 showed a good fit to the average daily MIR spectrum at day 1.5 (R^2cv = 0.79), and corresponding R^2 (0.87) of calibration was higher than that (0.76) obtained from FA profiles and CH_4 emissions (Dehareng et al., 2012). MIR spectra combined with lactation stage information can be used to simulate the variation in CH_4 emissions from cows as the lactation changes (the residual value is smaller), where R^2 and standard error of calibration equations were 0.75, and 63 g/d (Vanlierde et al., 2015).

8. Research Prospects

As a member of the spectroscopy family, IR has been significantly developed in recent decades. Compared with reference chemical methods, it has the following advantages:

1) Fast non-destructive testing: It is suitable for real-time on-line monitoring of milk, and it can provide information on milk components and the physiological state of dairy cows.
2) Detection in terms of population: The rapid analysis of spectroscopic techniques enables it to provide analysis of a large number of samples over a short period, so that producers can have production information on various farm aspects or dairy cow populations and make adjustments in time. In combination with genetic analysis, IR also can be used for trait screening at the population scale, which greatly facilitates breeding work.
3) Simultaneous detection of multiple traits: The infrared spectrum can simultaneously reflect the information of more than one trait due to the infrared absorption response with overtone and harmonic vibrations of many chemical bonds. This is different from traditional chemical analysis which can only measure one trait at a time.

However, as a methodology, the application of IR prediction can be affected by many factors, such as the spectroscopy instruments used, database construction, and chemometric methods. The limitations are mainly reflected in the following aspects:

1) Spectral analysis techniques and instrumentation: There are many overlaps in characteristic absorption regions for different chemical groups in the MIR fingerprint and NIR region. Although FT provides a means of deconvolution, the spectral complexity still needs to be overcome in the analysis. Variation of the light source under different environmental conditions will cause different degrees of spectral shift, so it is necessary to continuously calibrate the optical instrument. Aging of various parts of optical equipment in the instrument will decrease signal/noise ratio and affect the performance of the instrument.
2) Database construction: The determination of reference values is critical to the reliability of the database. Sampling and reference methods are directly related to the accuracy of spectra prediction (McParland et al., 2012; De Marchi et al., 2014). Expression of results with different units may indirectly affect the spectra model (Soyeurt et al., 2006; Rutten et al., 2009). The database size is also directly related to the prediction accuracy of the model (Rutten et al., 2010). Compatibility between databases (Grelet et al., 2015) also greatly influences the application of IR methods and the comparison between models.
3) The mathematics algorithm is embodied in three aspects: data pre-processing, region selection, and equation development. The suitable application for each spectral pretreatment should be seriously considered. The selection of responding wavenumbers or wavelengths based on known characteristic absorption regions of chemical components can largely avoid the interference of noise from other components, but for the development of new trait predictions corresponding characteristic regions have to be explored. There are many algorithms for establishing the relationship between spectral data and reference values. The equations obtained by the advanced algorithm are better than the general algorithm (Balabin and Smirnov, 2011), but the complexity of an advanced algorithm is not conducive to the promotion of this methodology.

9. Conclusion

The development of IR provides a quick and easy way to analyse milk components. Through a close combination with chemometrics, IR has been widely used in high-throughput non-destructive testing of products. In the future IR for milk testing may be developed in the following aspects: 1) IR has the potential to predict mineral and active substances in milk, so combined with appropriate laboratory analysis methods and optimised chemometrics, prediction of micro-nutrients in milk could be realised, and further indications of milk quality changes are expected; 2) Different reference methods and expression units have a great impact on the IR prediction, which would hinder comparison between the prediction results from different laboratory IR models and the promotion of this method, it is therefore necessary to standardise the detection method and the expression unit; 3) The spectra can also be used to characterise milk,

as some regions are heritable and some change with the environment, which can be used for monitoring the environmental characteristics of dairy cattle feeding or diet features of dairy animals; 4) IR can also be analysed simultaneously with disease diagnosis and reproductive traits for dairy cows, to find potential spectral changes in the pathological state of cows and genital organ activity, providing indicators for rapid diagnosis of disease and early detection of reproductive disorders. All of these require the continuous improvement of the precision of spectroscopy instruments, optimisation of the corresponding mathematical algorithms, and improvement of the reference methods, with the aim of improving the stability and reliability of the spectral model.

10. Acknowledgements

This study was financially supported by Agro-scientific Research in the Public Interest (201403071), the Agricultural Science and Technology Innovation Program (ASTIP-IAS12), and the Special Fund for Management of Modern Agricultural Industry Technology Systems (nycytx-04-01) granted by the Ministry of Agriculture (PRC).

11. References

Aernouts, B., E. Polshin, J. Lammertyn, and W. Saeys. 2011a. Visible and near-infrared spectroscopic analysis of raw milk for cow health monitoring: Reflectance or transmittance? Journal of Dairy Science 94(11):5315-5329.

Aernouts, B., E. Polshin, W. Saeys, and J. Lammertyn. 2011b. Mid-infrared spectrometry of milk for dairy metabolomics: A comparison of two sampling techniques and effect of homogenization. Anal Chim Acta 705(1-2):88-97.

Al-Qadiri, H. M., M. Lin, M. A. Al-Holy, A. G. Cavinato, and B. A. Rasco. 2008. Monitoring Quality Loss of Pasteurized Skim Milk Using Visible and Short Wavelength Near-Infrared Spectroscopy and Multivariate Analysis. Journal of Dairy Science 91(3):950-958.

Aliakbarian, B., M. Casale, M. Paini, A. A. Casazza, S. Lanteri, and P. Perego. 2015. Production of a novel fermented milk fortified with natural antioxidants and its analysis by NIR spectroscopy. LWT - Food Science and Technology 62(1):376-383.

Arbuckle, G. A., N. M. Buecheler, J. W. Hall, K. G. Valentine, S. Lefrant, J. Y. Mevellec, and E. Mulazzi. 1996. Spectroscopic properties of polyacetylenes synthesized via three modifications of Ziegler-Natta catalytic system. Synthetic Met 79(3):183-188.

Balabin, R. M. and S. V. Smirnov. 2011. Melamine detection by mid- and near-infrared (MIR/NIR) spectroscopy: A quick and sensitive method for dairy products analysis including liquid milk, infant formula, and milk powder. Talanta 85(1):562-568.

Barbano, D. M., K. L. Wojciechowski, and J. M. Lynch. 2010. Effect of preservatives

on the accuracy of mid-infrared milk component testing. Journal of Dairy Science 93(12):6000-6011.
Bastin, C., D. P. Berry, H. Soyeurt, and N. Gengler. 2012. Genetic correlations of days open with production traits and contents in milk of major fatty acids predicted by mid-infrared spectrometry. Journal Of Dairy Science 95(10):6113-6121.
Berget, I., H. Martens, A. Kohler, S. K. Sjurseth, N. K. Afseth, B. Narum, T. Adnoy, and S. Lien. 2010. Caprine CSN1S1 haplotype effect on gene expression and milk composition measured by Fourier transform infrared spectroscopy. Journal of Dairy Science 93(9):4340-4350.
Biggs, D. A. and D. McKenna. 1989. Alternative methods for infrared analysis of fat in milk: interlaboratory study. Journal - Association of Official Analytical Chemists 72(5):724-734.
Bittante, G. and A. Cecchinato. 2013. Genetic analysis of the Fourier-transform infrared spectra of bovine milk with emphasis on individual wavelengths related to specific chemical bonds. Journal of Dairy Science 96(9):5991-6006.
Bittante, G., A. Ferragina, C. Cipolat-Gotet, and A. Cecchinato. 2014. Comparison between genetic parameters of cheese yield and nutrient recovery or whey loss traits measured from individual model cheese-making methods or predicted from unprocessed bovine milk samples using Fourier-transform infrared spectroscopy. Journal Of Dairy Science 97(10):6560-6572.
Bogomolov, A. and M. Hachey. 2007. Application of SIMPLISMA purity function for variable selection in multivariate regression analysis: A case study of protein secondary structure determination from infrared spectra. Chemometr Intell Lab 88(1):132-142.
Bonfatti, V., G. Di Martino, and P. Carnier. 2011. Effectiveness of mid-infrared spectroscopy for the prediction of detailed protein composition and contents of protein genetic variants of individual milk of Simmental cows. Journal Of Dairy Science 94(12):5776-5785.
Borin, A., M. F. Ferrao, C. Mello, D. A. Maretto, and R. J. Poppi. 2006. Least-squares support vector machines and near infrared spectroscopy for quantification of common adulterants in powdered milk. Anal Chim Acta 579(1):25-32.
Botelho, B. G., N. Reis, L. S. Oliveira, and M. M. Sena. 2015. Development and analytical validation of a screening method for simultaneous detection of five adulterants in raw milk using mid-infrared spectroscopy and PLS-DA. Food Chem 181:31-37.
Calamari, L., L. Gobbi, and P. Bani. 2016. Improving the prediction ability of FT-MIR spectroscopy to assess titratable acidity in cow's milk. Food Chem 192:477-484.
Capuano, E., J. Rademaker, H. van den Bijgaart, and S. M. van Ruth. 2014. Verification of fresh grass feeding, pasture grazing and organic farming by FTIR spectroscopy analysis of bovine milk. Food Res Int 60:59-65.
Cassoli, L. D., B. Sartori, and P. F. Machado. 2011. The use of the Fourier Transform Infrared spectroscopy to determine adulterants in raw milk. Revista Brasileira De

Zootecnia-Brazilian Journal of Animal Science 40(11):2591-2596.
Cecchinato, A., C. Cipolat-Gotet, J. Casellas, M. Penasa, A. Rossoni, and G. Bittante. 2013. Genetic analysis of rennet coagulation time, curd-firming rate, and curd firmness assessed over an extended testing period using mechanical and near-infrared instruments. Journal of Dairy Science 96(1):50-62.
Cecchinato, A., M. De Marchi, L. Gallo, G. Bittante, and P. Carnier. 2009. Mid-infrared spectroscopy predictions as indicator traits in breeding programs for enhanced coagulation properties of milk. Journal of Dairy Science 92(10):5304-5313.
Chessa, S., O. Bulgari, R. Rizzi, L. Calamari, P. Bani, S. Biffani, and A. M. Caroli. 2014. Selection for milk coagulation properties predicted by Fourier transform infrared spectroscopy in the Italian Holstein-Friesian breed. Journal of Dairy Science 97(7):4512-4521.
Chu, Y. 2012. Study on Discrimination of Fresh of Goat's Milk and Adulteration Goat's Milk by Near Infrared Spectroscopy Technology. Northwest A&F University, Yangling.
Coppa, M., A. Ferlay, C. Leroux, M. Jestin, Y. Chilliard, B. Martin, and D. Andueza. 2010. Prediction of milk fatty acid composition by near infrared reflectance spectroscopy. Int Dairy J 20(3):182-189.
Coppa, M., B. Martin, C. Agabriel, C. Chassaing, C. Sibra, I. Constant, B. Graulet, and D. Andueza. 2012. Authentication of cow feeding and geographic origin on milk using visible and near-infrared spectroscopy. Journal of Dairy Science 95(10):5544-5551.
Coppa, M., A. Revello-Chion, D. Giaccone, A. Ferlay, E. Tabacco, and G. Borreani. 2014. Comparison of near and medium infrared spectroscopy to predict fatty acid composition on fresh and thawed milk. Food Chem 150:49-56.
Dagnachew, B. S., A. Kohler, and T. Adnoy. 2013a. Genetic and environmental information in goat milk Fourier transform infrared spectra. Journal Of Dairy Science 96(6):3973-3985.
Dagnachew, B. S., T. H. E. Meuwissen, and T. Adnoy. 2013b. Genetic components of milk Fourier-transform infrared spectra used to predict breeding values for milk composition and quality traits in dairy goats. Journal Of Dairy Science 96(9):5933-5942.
Dal Zotto, R., M. De Marchi, A. Cecchinato, M. Penasa, M. Cassandro, P. Carnier, L. Gallo, and G. Bittante. 2008. Reproducibility and Repeatability of Measures of Milk Coagulation Properties and Predictive Ability of Mid-Infrared Reflectance Spectroscopy. Journal of Dairy Science 91(10):4103-4112.
De Marchi, M., V. Bonfatti, A. Cecchinato, G. Di Martino, and P. Carnier. 2009a. Prediction of protein composition of individual cow milk using mid-infrared spectroscopy. Ital J Anim Sci 8:399-401.
De Marchi, M., C. C. Fagan, C. P. O'Donnell, A. Cecchinato, R. Dal Zotto, M. Cassandro, M. Penasa, and G. Bittante. 2009b. Prediction of coagulation properties, titratable acidity, and pH of bovine milk using mid-infrared spectroscopy. Journal of

Dairy Science 92(1):423-432.
De Marchi, M., M. Penasa, A. Cecchinato, M. Mele, P. Secchiari, and G. Bittante. 2011. Effectiveness of mid-infrared spectroscopy to predict fatty acid composition of Brown Swiss bovine milk. Animal 5(10):1653-1658.
De Marchi, M., V. Toffanin, M. Cassandro, and M. Penasa. 2013. Prediction of coagulating and noncoagulating milk samples using mid-infrared spectroscopy. Journal Of Dairy Science 96(7):4707-4715.
De Marchi, M., V. Toffanin, M. Cassandro, and M. Penasa. 2014. Invited review: Mid-infrared spectroscopy as phenotyping tool for milk traits. Journal of Dairy Science 97(3):1171-1186.
de Roos, A. P. W., H. J. C. M. van den Bijgaart, J. Hørlyk, and G. de Jong. 2007. Screening for Subclinical Ketosis in Dairy Cattle by Fourier Transform Infrared Spectrometry. Journal of Dairy Science 90(4):1761-1766.
Dehareng, F., C. Delfosse, E. Froidmont, H. Soyeurt, C. Martin, N. Gengler, A. Vanlierde, and P. Dardenne. 2012. Potential use of milk mid-infrared spectra to predict individual methane emission of dairy cows. Animal 6(10):1694-1701.
Domingo, E., A. A. Tirelli, C. A. Nunes, M. C. Guerreiro, and S. M. Pinto. 2014. Melamine detection in milk using vibrational spectroscopy and chemometrics analysis: A review. Food Res Int 60:131-139.
Etzion, Y., R. Linker, U. Cogan, and I. Shmulevich. 2004. Determination of Protein Concentration in Raw Milk by Mid-Infrared Fourier Transform Infrared/Attenuated Total Reflectance Spectroscopy. Journal of Dairy Science 87(9):2779-2788.
Ferragina, A., C. Cipolat-Gotet, A. Cecchinato, and G. Bittante. 2013. The use of Fourier-transform infrared spectroscopy to predict cheese yield and nutrient recovery or whey loss traits from unprocessed bovine milk samples. Journal of Dairy Science 96(12):7980-7990.
Ferrand-Calmels, M., I. Palhière, M. Brochard, O. Leray, J. M. Astruc, M. R. Aurel, S. Barbey, F. Bouvier, P. Brunschwig, H. Caillat, M. Douguet, F. Faucon-Lahalle, M. Gelé, G. Thomas, J. M. Trommenschlager, and H. Larroque. 2014. Prediction of fatty acid profiles in cow, ewe, and goat milk by mid-infrared spectrometry. Journal of Dairy Science 97(1):17-35.
Ferrand, M., B. Huquet, S. Barbey, F. Barillet, F. Faucon, H. Larroque, O. Leray, J. M. Trommenschlager, and M. Brochard. 2011. Determination of fatty acid profile in cow's milk using mid-infrared spectrometry: Interest of applying a variable selection by genetic algorithms before a PLS regression. Chemometr Intell Lab 106(2):183-189.
Grelet, C., J. A. Fernández Pierna, P. Dardenne, V. Baeten, and F. Dehareng. 2015. Standardization of milk mid-infrared spectra from a European dairy network. Journal of Dairy Science 98(4):2150-2160.
Gustavsson, F., M. Glantz, N. A. Poulsen, L. Wadsö, H. Stålhammar, A. Andrén, H. Lindmark Månsson, L. B. Larsen, M. Paulsson, and W. F. Fikse. 2014. Genetic parameters for rennet- and acid-induced coagulation properties in milk from Swedish Red dairy cows. Journal of Dairy Science 97(8):5219-5229.

Hansen, P. W. 1999. Screening of Dairy Cows for Ketosis by Use of Infrared Spectroscopy and Multivariate Calibration. Journal of Dairy Science 82(9):2005-2010.
Heuer, C., H. J. Luinge, E. T. G. Lutz, Y. H. Schukken, J. H. van der Maas, H. Wilmink, and J. P. T. M. Noordhuizen. 2001. Determination of Acetone in Cow Milk by Fourier Transform Infrared Spectroscopy for the Detection of Subclinical Ketosis. Journal of Dairy Science 84(3):575-582.
Huang, Y., S. Min, J. Duan, L. Wu, and Q. Li. 2014. Identification of additive components in powdered milk by NIR imaging methods. Food Chem 145(0):278-283.
Huang, Y., K. Tian, S. Min, Y. Xiong, and G. Du. 2015. Distribution assessment and quantification of counterfeit melamine in powdered milk by NIR imaging methods. Food Chem 177(0):174-181.
Inácio, M. R. C., M. d. F. V. de Moura, and K. M. G. de Lima. 2011. Classification and determination of total protein in milk powder using near infrared reflectance spectrometry and the successive projections algorithm for variable selection. Vibrational Spectroscopy 57(2):342-345.
Jaiswal, P., S. N. Jha, A. Borah, A. Gautam, M. K. Grewal, and G. Jindal. 2015. Detection and quantification of soymilk in cow–buffalo milk using Attenuated Total Reflectance Fourier Transform Infrared spectroscopy (ATR–FTIR). Food Chem 168(0):41-47.
Jawaid, S., F. N. Talpur, S. T. H. Sherazi, S. M. Nizamani, and A. A. Khaskheli. 2013. Rapid detection of melamine adulteration in dairy milk by SB-ATR-Fourier transform infrared spectroscopy. Food Chem 141(3):3066-3071.
Jin, Y., B. Du, and X. Zhi. 2016. Fast recognition of adulterated milk and the brands of milk based on the NIR technology. Food Research And Development 37(3):178-181.
Kawasaki, M., S. Kawamura, M. Tsukahara, S. Morita, M. Komiya, and M. Natsuga. 2008. Near-infrared spectroscopic sensing system for on-line milk quality assessment in a milking robot. Computers and Electronics in Agriculture 63(1):22-27.
Kaylegian, K. E., D. A. Dwyer, J. M. Lynch, D. E. Bauman, J. R. Fleming, and D. M. Barbano. 2009. Impact of fatty acid composition on the accuracy of mid-infrared fat analysis of farm milks. Journal of Dairy Science 92(6):2502-2513.
Kong, W. W., C. Zhang, F. Liu, A. P. Gong, and Y. He. 2013. Irradiation dose detection of irradiated milk powder using visible and near-infrared spectroscopy and chemometrics. Journal of Dairy Science 96(8):4921-4927.
Leclercq, G., N. Gengler, H. Soyeurt, and C. Bastin. 2013. Genetic variability of the mid-infrared prediction of lactoferrin content in milk for Walloon Holstein first-parity cows. Livest Sci 151(2-3):158-162.
Li, L. and W. Ding. 2010. Discriminant analysis of raw milk adulterated with botanical filling material using near infrared spectroscopy. Spectroscopy and Spectral Analysis 30(5):1238-1242.
Liu, J., J. Ren, Z. M. Liu, and B. H. Guo. 2015. A new comprehensive index for discriminating adulteration in bovine raw milk. Food Chem 172:251-256.

Luginbuhl, W. 2002. Evaluation of designed calibration samples for casein calibration in Fourier transform infrared analysis of milk. Lebensm-Wiss Technol 35(6):554-558.
Lyndgaard, C. B., S. B. Engelsen, and F. W. J. van den Berg. 2012. Real-time modeling of milk coagulation using in-line near infrared spectroscopy. Journal of Food Engineering 108(2):345-352.
Maâmouri, O., H. Rouissi, S. Dridi, M. Kammoun, J. De Baerdemaeker, and R. Karoui. 2008. Mid infrared attenuated total reflection spectroscopy as a rapid tool to assess the quality of Sicilo-Sarde ewe's milk during the lactation period after replacing soybean meal with scotch bean in the feed ration. Food Chem 106(1):361-368.
Markovich, R. J. and C. Pidgeon. 1991. Introduction to Fourier transform infrared spectroscopy and applications in the pharmaceutical sciences. Pharmaceutical research 8(6):663-675.
Maroteau, C., I. Palhiere, H. Larroque, V. Clement, M. Ferrand, G. Tosser-Klopp, and R. Rupp. 2014. Genetic parameter estimation for major milk fatty acids in Alpine and Saanen primiparous goats. Journal Of Dairy Science 97(5):3142-3155.
Maurice-Van Eijndhoven, M. H., H. Bovenhuis, H. Soyeurt, and M. P. Calus. 2013. Differences in milk fat composition predicted by mid-infrared spectrometry among dairy cattle breeds in the Netherlands. Journal of Dairy Science 96(4):2570-2582.
McParland, S., G. Banos, B. McCarthy, E. Lewis, M. P. Coffey, B. O'Neill, M. O'Donovan, E. Wall, and D. P. Berry. 2012. Validation of mid-infrared spectrometry in milk for predicting body energy status in Holstein-Friesian cows. Journal of Dairy Science 95(12):7225-7235.
McParland, S., E. Kennedy, E. Lewis, S. G. Moore, B. McCarthy, M. O'Donovan, and D. P. Berry. 2015. Genetic parameters of dairy cow energy intake and body energy status predicted using mid-infrared spectrometry of milk. Journal of Dairy Science 98(2):1310-1320.
McParland, S., E. Lewis, E. Kennedy, S. G. Moore, B. McCarthy, M. O'Donovan, S. T. Butler, J. E. Pryce, and D. P. Berry. 2014. Mid-infrared spectrometry of milk as a predictor of energy intake and efficiency in lactating dairy cows. Journal of Dairy Science 97(9):5863-5871.
Melfsen, A., E. Hartung, and A. Haeussermann. 2012a. Accuracy of in-line milk composition analysis with diffuse reflectance near-infrared spectroscopy. Journal of Dairy Science 95(11):6465-6476.
Melfsen, A., E. Hartung, and A. Haeussermann. 2012b. Accuracy of milk composition analysis with near infrared spectroscopy in diffuse reflection mode. Biosystems Engineering 112(3):210-217.
Mouazen, A. M., S. Dridi, H. Rouissi, J. De Baerdemaeker, and H. Ramon. 2009. Prediction of selected ewe's milk properties and differentiating between pasture and box feeding using visible and near infrared spectroscopy. Biosystems Engineering 104(3):353-361.
Nicolaou, N., Y. Xu, and R. Goodacre. 2010. Fourier transform infrared spectroscopy and multivariate analysis for the detection and quantification of different milk species.

Journal of Dairy Science 93(12):5651-5660.
Pappas, C. S., P. A. Tarantilis, E. Moschopoulou, G. Moatsou, I. Kandarakis, and M. G. Polissiou. 2008. Identification and differentiation of goat and sheep milk based on diffuse reflectance infrared Fourier transform spectroscopy (DRIFTS) using cluster analysis. Food Chem 106(3):1271-1277.
Paré, J. R. J. and J. M. R. Bélanger. 1997. Instrumental methods in food analysis. Techniques and instrumentation in analytical chemistry. No. 18. Elsevier, Amsterdam ; New York.
Penasa, M., M. De Marchi, S. Ton, L. Ancilotto, and M. Cassandro. 2015. Reproducibility and repeatability of milk coagulation properties predicted by mid-infrared spectroscopy. Int Dairy J 47:1-5.
Purnomoadi, A., K. K. Batajoo, K. Ueda, and F. Terada. 1999. Influence of feed source on determination of fat and protein in milk by near-infrared spectroscopy. Int Dairy J 9(7):447-452.
Rivero, M. J., X. Valderrama, D. Haines, and D. Alomar. 2012. Prediction of immunoglobulin G content in bovine colostrum by near-infrared spectroscopy. Journal of Dairy Science 95(3):1410-1418.
Rodriguez-Saona, L. E. and M. E. Allendorf. 2011. Use of FTIR for rapid authentication and detection of adulteration of food. Annu Rev Food Sci T 2:467-483.
Roggo, Y., P. Chalus, L. Maurer, C. Lema-Martinez, A. Edmond, and N. Jent. 2007. A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies. J Pharmaceut Biomed 44(3):683-700.
Rutten, M. J. M., H. Bovenhuis, J. M. L. Heck, and J. A. M. van Arendonk. 2011a. Predicting bovine milk protein composition based on Fourier transform infrared spectra. Journal Of Dairy Science 94(11):5683-5690.
Rutten, M. J. M., H. Bovenhuis, J. M. L. Heck, and J. A. M. van Arendonk. 2011b. Prediction of beta-lactoglobulin genotypes based on milk Fourier transform infrared spectra. Journal Of Dairy Science 94(8):4183-4188.
Rutten, M. J. M., H. Bovenhuis, K. A. Hettinga, H. J. F. van Valenberg, and J. A. M. van Arendonk. 2009. Predicting bovine milk fat composition using infrared spectroscopy based on milk samples collected in winter and summer. Journal of Dairy Science 92(12):6202-6209.
Rutten, M. J. M., H. Bovenhuis, and J. A. M. van Arendonk. 2010. The effect of the number of observations used for Fourier transform infrared model calibration for bovine milk fat composition on the estimated genetic parameters of the predicted data. Journal of Dairy Science 93(10):4872-4882.
Santos, P. M., E. R. Pereira-Filho, and L. E. Rodriguez-Saona. 2013a. Rapid detection and quantification of milk adulteration using infrared microspectroscopy and chemometrics analysis. Food Chem 138(1):19-24.
Santos, P. M., E. R. Pereira, and L. E. Rodriguez-Saona. 2013b. Application of hand-held and portable infrared spectrometers in bovine milk analysis. J Agr Food Chem 61(6):1205-1211.

Shi, Y. 2014. Identification Study of Milk Quality and Metamorphic Grade Based Spectroscopic Techniques. Chongqing University of Technology, Chongqing.
Sivakesava, S. and J. Irudayaraj. 2002. Rapid Determination of Tetracycline in Milk by FT-MIR and FT-NIR Spectroscopy. Journal of Dairy Science 85(3):487-493.
Soyeurt, H., C. Bastin, F. G. Colinet, V. M. R. Arnould, D. P. Berry, E. Wall, F. Dehareng, H. N. Nguyen, P. Dardenne, J. Schefers, J. Vandenplas, K. Weigel, M. Coffey, L. Theron, J. Detilleux, E. Reding, N. Gengler, and S. McParland. 2012. Mid-infrared prediction of lactoferrin content in bovine milk: potential indicator of mastitis. Animal 6(11):1830-1838.
Soyeurt, H., D. Bruwier, J. M. Romnee, N. Gengler, C. Bertozzi, D. Veselko, and P. Dardenne. 2009. Potential estimation of major mineral contents in cow milk using mid-infrared spectrometry. Journal of Dairy Science 92(6):2444-2454.
Soyeurt, H., F. G. Colinet, V. M. R. Arnould, P. Dardenne, C. Bertozzi, R. Renaville, D. Portetelle, and N. Gengler. 2007. Genetic variability of lactoferrin content estimated by mid-infrared Spectrometry in bovine milk. Journal Of Dairy Science 90(9):4443-4450.
Soyeurt, H., P. Dardenne, F. Dehareng, G. Lognay, D. Veselko, M. Marlier, C. Bertozzi, P. Mayeres, and N. Gengler. 2006. Estimating fatty acid content in cow milk using mid-infrared spectrometry. Journal of Dairy Science 89(9):3690-3695.
Soyeurt, H., F. Dehareng, N. Gengler, S. McParland, E. Wall, D. P. Berry, M. Coffey, and P. Dardenne. 2011. Mid-infrared prediction of bovine milk fatty acids across multiple breeds, production systems, and countries. Journal of Dairy Science 94(4):1657-1667.
Soyeurt, H., I. Misztal, and N. Gengler. 2010. Genetic variability of milk components based on mid-infrared spectral data. Journal of Dairy Science 93(4):1722-1728.
Toffanin, V., M. De Marchi, N. Lopez-Villalobos, and M. Cassandro. 2015. Effectiveness of mid-infrared spectroscopy for prediction of the contents of calcium and phosphorus, and titratable acidity of milk and their relationship with milk quality and coagulation properties. Int Dairy J 41:68-73.
Tsenkova, R., S. Atanassova, Y. Ozaki, K. Toyoda, and K. Itoh. 2001. Near-infrared spectroscopy for biomonitoring: influence of somatic cell count on cow's milk composition analysis. Int Dairy J 11(10):779-783.
Tsenkova, R., S. Atanassova, K. Toyoda, Y. Ozaki, K. Itoh, and T. Fearn. 1999. Near-Infrared Spectroscopy for Dairy Management: Measurement of Unhomogenized Milk Composition. Journal of Dairy Science 82(11):2344-2351.
Valenti, B., B. Martin, D. Andueza, C. Leroux, C. Labonne, F. Lahalle, H. Larroque, P. Brunschwig, C. Lecomte, M. Brochard, and A. Ferlay. 2013. Infrared spectroscopic methods for the discrimination of cows' milk according to the feeding system, cow breed and altitude of the dairy farm. Int Dairy J 32(1):26-32.
van Knegsel, A. T. M., S. G. A. van der Drift, M. Horneman, A. P. W. de Roos, B. Kemp, and E. A. M. Graat. 2010. Short communication: Ketone body concentration in milk determined by Fourier transform infrared spectroscopy: Value for the

detection of hyperketonemia in dairy cows. Journal of Dairy Science 93(7):3065-3069.
Vanlierde, A., M. L. Vanrobays, F. Dehareng, E. Froidmont, H. Soyeurt, S. McParland, E. Lewis, M. H. Deighton, F. Grandl, M. Kreuzer, B. Gredler, P. Dardenne, and N. Gengler. 2015. Hot topic: Innovative lactation-stage-dependent prediction of methane emissions from milk mid-infrared spectra. Journal of Dairy Science 98(8):5740-5747.
Visentin, G., A. McDermott, S. McParland, D. P. Berry, O. A. Kenny, A. Brodkorb, M. A. Fenelon, and M. De Marchi. 2015. Prediction of bovine milk technological traits from mid-infrared spectroscopy analysis in dairy cows. J Dairy Sci 98(9):6620-6629.
Wang, M., F. Yu, and X. Liu. 2015. Detection of protein, fat in milk by using near infrared diffuse reflectance technique. Laser Journal 36(1):70-73.
Wu, D., Y. He, and S. Feng. 2008. Short-wave near-infrared spectroscopy analysis of major compounds in milk powder and wavelength assignment. Anal Chim Acta 610(2):232-242.
Yang, J., D. Bu, and J. Wang. 2013. Determination of chemical composition of milk using near-infrared reflectance spectroscopy. Food Science 34(20):153-156.
Yang, L. 2011. Study of Infrared Model on Rapid Detection of Five Fatty Acids in Milk Fat. Shandong University of Technology, Zibo.
Yang, R., R. Liu, and K. Xu. 2012. Detection of urea in milk using two-dimensional correlation spectroscopy and partial least square method. Transactions of the Chinese Society of Agricultural Engineering 28(6):259-263.
Yazdanpanah, N. and T. A. G. Langrish. 2013. Comparative study of deteriorative changes in the ageing of milk powder. Journal of Food Engineering 114(1):14-21.
Zhang, H. 2010. Rapid Detection of Milk Protein and Fat Content Using Near Infrared Spectroscopy. Vol. PhD. Central South University, Changsha.
Zhang, L.-G., X. Zhang, L.-J. Ni, Z.-B. Xue, X. Gu, and S.-X. Huang. 2014. Rapid identification of adulterated cow milk by non-linear pattern recognition methods based on near infrared spectroscopy. Food Chem 145:342-348.
Zhao, M., Y. Shi, and B. Luo. 2014. Testing the amount of protein and fat contained in the milk based on near in frared transmission spectroscopy. Laser Journal 35(1):44-50.
Zou, X. B., J. W. Zhao, M. J. W. Povey, M. Holmes, and H. P. Mao. 2010. Variables selection methods in near-infrared spectroscopy. Anal Chim Acta 667(1-2):14-32.

Doctoral Trainings and Scientific Communications

1.	**Thematic Training** (minimum 15 credits)	Credits
1.1	RDCT0369-1 - Training session: Vibrational Spectroscopy and Chemometrics	4
1.2	5th international symposium on dairy cow nutrition and milk quality	3
1.3	Training of proteomics and identification of unknown low weight molecular using mass spectrometry	1
1.4	The 2016 World Life Science Conference	3
1.5	RDCT0273-1 - Advances on animal nutrition and feed science	1
1.6	Detection technology exchange meeting on goat/sheep milk and milk product identification	1
1.7	PhD student's seminar	3
1.8	Quality Control in Food Microbiology Laboratory	2
1.9	4th international symposium on dairy cow nutrition and milk quality	2
	Subtotal	20
2.	**Transversal Training** (minimum 10 credits)	Credits
2.1	Comprehensive examination of PhD student	3
2.2	RDCT0274-1 - Literature learning	2
2.3	Regulation training: expert interpretation of "Food Safety Law in Chinese 2015"	1
2.4	RDCT0270-1 - Chinese Marxism and Contemporary	2
2.5	RDCT0271-1 - English (PhD)	2
2.6	RDCT0272-1 - Lectures on Agricultural Science and Technology Progress	2
	Subtotal	12
3.	**Scientific Communications** (minimum 25 credits)	Credits
3.1	Detection of plant protein in adulterated milk using nontargeted nano-high- performance liquid chromatography- tandem mass spectroscopy combined with principal component analysis	8
3.2	Detection of plant protection adulterated in fluid milk using two-dimensional gel electrophoresis combined with mass spectrometry	8
3.3	Comparative milk fatty acid analysis of different dairy species	8
3.4	Detection of plant protein in adulterated milk by SDS-PAGE and LC-MS-experiment progress	3
3.5	The effect of Heat Treatment on Milk Components and Heat-Sensitive Indicator Variation	8
3.6	Research advances on infrared spectrum in milk production and determination	8
3.7	Detection of plant protein and fat adulteration in milk	3
	Subtotal	46
	In total	78