Applications of front-face fluorescence spectroscopy and chemometrics to measure casein content in milk and detect protein leaks in dairy ultrafiltration permeates

by

Yizhou Ma

B.S., University of Minnesota, 2017

A THESIS

submitted in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

Food Science

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2019

Approved by:

Major Professor
Jayendra K. Amamcharla

Copyright

Abstract

Quantitative analysis of casein content in cheese milk can give a better control over cheese yield and understand cheese quality. Traditional analytical methods for casein measurement rely on nitrogen-based quantification and involve time-consuming sample preparation steps. The current study applied front-face fluorescence spectroscopy (FFFS) combined with chemometrics to quantify casein and casein-to-crude-protein ratio (CN/CP) in milk intended for cheese manufacturing. FFFS spectra of acid-precipitated casein milk dispersion (pH = 4.6) were collected from 20 ultrafiltered (UF) model milk samples with different casein contents. A preliminary calibration model was developed with principal component regression (PCR) using reference casein contents and the FFFS spectra. The model was externally validated with 20 raw milk samples and a root mean square error (RMSE) of 0.15% was found between the predicted and reference casein contents. A relative prediction error (RPE) of 6.7% indicating usefulness for quality control purposes. To further refine the FFFS-based casein quantification method, 30 model milk samples prepared from UF and microfiltration (MF) permeates and retentates to generate different casein contents and CN/CP. The FFFS spectra were collected following the same procedure and used as predictors for casein and CN/CP quantifications. Calibration models were developed using partial least squares regression (PLSR) and elastic net regression (ENR) and the models were further optimized using 20 samples including raw, skim, and UF milk. The optimized PLSR and ENR models were again tested using 20 test samples including raw, skim, and UF milk and evaluated in terms of RMSE, residual prediction deviation (RPD), and RPE. The PLSR and ENR models reduced the RMSE for casein quantification to 0.13% with RPD ranged from 3.2 to 3.4, indicating practical model performances. For CN/CP quantification, PLSR models resulted in useful predictions with an RMSE of 0.024, an RPD of

1.5, and an RPE of 3.0%. The FFFS-based casein quantification method provides a rapid casein measurement in fluid milk and can be implemented in the cheese industry for routine measurements.

In a different study, FFFS was utilized to predict the protein leaks in permeate during membrane processing of skim milk and whey. Protein leak occurs when proteinous matters pass through the UF membrane into the permeate stream leading to financial losses and product quality defects. FFFS as a sensitive and specific instrument was applied to characterize protein leak occurrences in UF permeate, develop chemometrics models to quantify true protein (TP) content in permeate streams, and classify sources of protein leak in the feed material. Measurements of crude protein (CP), non-protein nitrogen, TP, tryptone-equivalent peptide, α-lactalbumin (α-LA), and β-lactoglobulin (β-LG) were performed on 33 lots of commercial whey permeate and 29 lots of commercial milk permeate. Protein leaks were attributed to high TP, high-peptide, and presence of α-LA or β-LG. Tryptophan was identified as the fluorophore of interest for protein leak detection based on the excitation-emission matrix analysis of representative permeate with high and low TP contents. Quantitative models based on PLSR were developed using tryptophan excitation spectra and true protein content in the permeate. The model yielded a RMSE of 0.22% (dry-basis) and RPD of 2.8 based on external validations, showing a useful model for quality control purposes. Moreover, classification models based on partial least squares discriminant analysis were developed to detect high TP level, high peptide level, and presence of α-LA or β-LG with 83.3%, 84.8%, and 98.5% cross-validated accuracy, respectively. This method showed that FFFS and chemometrics can rapidly detect protein leak and identify the source of protein leak in UF permeate, which can reduce financial loss from protein leak and maintain high-quality permeate production.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to express my sincere gratitude to my major advisor, Dr. Jayendra Amamcharla, for his guidance, care, and motivation throughout my master's program. I would like to thank Drs. Karen Schmidt and Yonghui Li for serving on my committee. I appreciate the technical advice and instrumental support from Dr. Inès-Birlouez-Aragon, managing director from Spectralys Innovation (Romainville, France). I would like to extend my thank to Ms. Kathy Lowe from the Soil Testing Lab of Kansas State University for technical assistance on protein measurements. I am thankful to my family, Food Science Institute faculty and staff, and other graduate students at Kansas State University for their supports. Finally, I would like to thank the Midwest Dairy Foods Research Center (St. Paul, MN) for their financial support.

Chapter 1 - Introduction

As the food industry continuously seeks innovative analytical methods, the spectroscopy-based techniques offer rapid and non-invasive analysis with promising possibilities in a wide range of applications. The spectroscopy instruments and chemometrics made it possible to provide on-line or at-line detection capability during food production in real-time. Food contains intrinsic fluorescence-emitting compounds (fluorophores) which can be used as compounds of interest in fluorescence spectroscopic analysis. In food science applications, front-face fluorescence spectroscopy (FFFS) has been applied due to its specificity, sensitivity, and ability to measure turbid samples (Andersen and Mortensen, 2008). The obtained spectra from FFFS measurement are analyzed by chemometrics, a field of study applying mathematical and statistical methods to extract chemical and physical information from complex data (Wold and Sjöström ,1998). Applications of FFFS have shown promising results to analyze dairy food composition, processing, and storage-induced changes.

Cheese is one of the most popular fermented dairy products in the United States with 12.7 billion pounds produced in 2017 (US Department of Agriculture, 2019). Casein content in milk has been identified as one of the yield indicators for cheese production (Emmons and Modler, 2010). Various cheese yield prediction equations include casein content in milk as a factor, and standardizing casein-to-fat ratio has been an industry practice to maintain consistency in the cheese production. In addition, casein content in cheese milk can influence the quality of finished product as it serves as the flavor substrate during aging and the texture backbone in the finished cheese (Fox, 1989). To maintain product consistency in cheese manufacturing, an accurate casein measurement in milk is necessary. FFFS combined with chemometrics can potentially fulfill such need from the cheese industry.

In addition to dairy products, dairy-based ingredients such as milk and whey protein concentrates (MPC and WPC, respectively) have gained market shares in the recent years (Agarwal et al., 2015). The production of MPC and WPC from skim milk or whey involves ultrafiltration (UF) and diafiltration to fractionate the protein-rich retentate from the lactose-rich permeate. The liquid permeate can then be evaporated, concentrated, crystalized, and spray-dried in powder form. Based on the original source of the permeate, they are branded as "milk permeate powder" and "deproteinized whey" and used in confectionary and bakery products. A small fraction of protein can leak through the membrane under certain conditions. The protein leak during UF can lead to financial losses and permeate powder deficiencies during storage. As a standard practice in the dairy industry, permeates are analyzed for total protein content only in the powder form resulting in a delay between the protein leak and detection. Moreover, measuring only total protein in the finished permeate powders does not reveal the source of protein leak, potentially coming from loss of true protein or peptides. FFFS combined with chemometrics can potentially offer real-time true protein analysis in liquid permeate to monitor protein leak during UF processing.

This thesis focuses on the development and validation of FFFS and chemometrics to measure casein in fluid milk and detect protein leak in UF permeates. The FFFS-based methods can be potentially implemented into dairy foods production facilities and maintain product qualities.

References

Agarwal, S., R. L. W. Beausire, S. Patel, and H. Patel. 2015. Innovative uses of milk protein concentrates in product development. J. Food Sci. 80:23-A29.

Andersen, C.M., and G. Mortensen. 2008. Fluorescence Spectroscopy: A Rapid Tool for Analyzing Dairy Products. J. Agric. Food Chem. 56:720–729.

Emmons, D.B., and H.W. Modler. 2010. Invited review: A commentary on predictive cheese yield formulas1. J. Dairy Sci. 93:5517–5537.

Fox, P.F. 1989. Proteolysis During Cheese Manufacture and Ripening1. J. Dairy Sci. 72:1379–1400.

US Department of Agriculture, National Agricultural Statistics Service. 2019. National statistics for cheese.

Wold, S., and M. Sjöström. 1998. Chemometrics, present and future success. Chemom. Intell. Lab. Syst. 44:3–14.

Chapter 2 - Literature Review

Milk composition

Bovine milk is defined as “the lacteal secretion, practically free from colostrum, obtained by the complete milking of one or more healthy cow” (21CFR131.110). Chemically speaking, bovine milk consists of water, fat, lactose, protein, and mineral substances. According to Walstra et al. (2005), the primary composition of milk is water (87% w/w), which serves as the solvent and continuous phase for other chemical compounds in milk. Among the non-water fractions of milk, lactose (4.6% w/w) as the distinctive disaccharide of glucose and galactose. Milk dry matter also contain fat (4% w/w), which mostly exists in the form of globules. The various fat compounds found in milk include triglycerides, phospholipids, cholesterol, free fatty acids, and mono and di-glycerides. Milk also contains various mineral substances such as potassium, sodium, calcium, magnesium, chloride, and phosphate. These minerals exist in both ionic and colloidal forms, contributing to the stabilization of the milk system. About 80% the proteins in milk (3.3% w/w) are caseins, with a combination of β-casein, κ-casein, αs_1-casein, and αs_2-casein. The reminder proteins are largely serum proteins including ß-lactoglobulin, α-lactalbumin, blood serum albumin, immunoglobulins, lactoferrin, transferrin, and many other proteins and enzymes.

This chapter offers reviews of peer-reviewed literature on milk casein, its structure, composition, and impacts on cheese manufacturing. The review then introduces various methods of measuring casein in milk using standard methods and rapid alternative methods. Front-face fluorescence spectroscopy (FFFS) is another focus of this chapter. Its instrumentation, working principle, and applications on food products are reviewed in the subsequent sections. Lastly, spectral analysis using chemometrics is discussed in this chapter as an essential step to develop accurate and reproducible spectroscopic rapid methods.

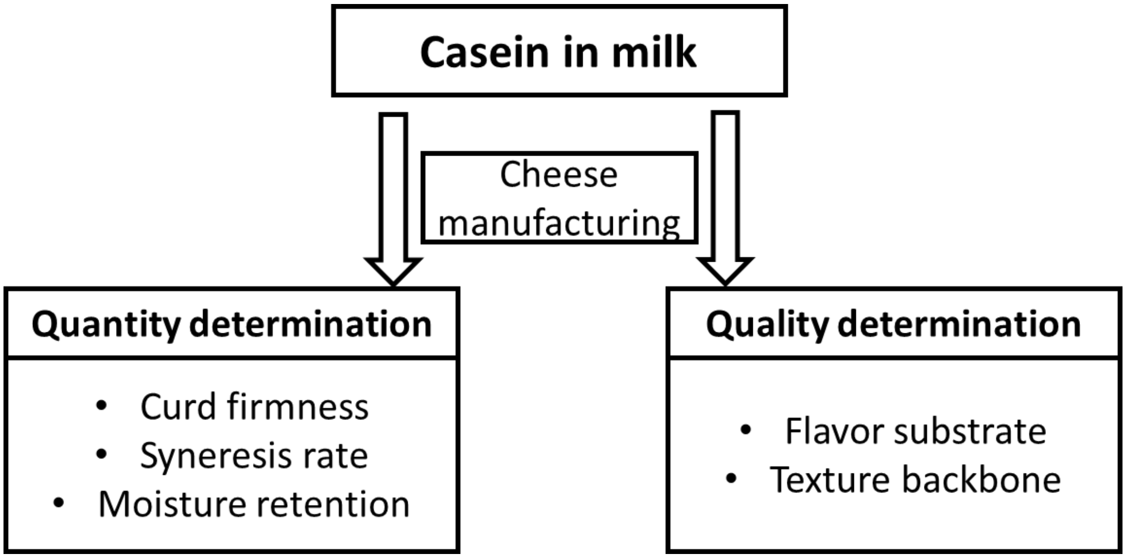
Casein chemistry

While milk as a complicated system relies on all compositions interacting harmoniously, food consumption trend has bene growing towards its protein fractions (Horne, 2006). Casein as the major milk protein has been studied extensively for its unique chemical structure. Casein exists in milk as micelles, voluminous, spherical, and negatively charged particles that hold water, casein proteins, and inorganic matters (calcium phosphate). Overall, micelle diameters were found from 100 to 300 nm with variations occurred based on milk compositions. Namely, the amount of κ-casein has been associated to affect casein micelle size, with more κ-casein contributing to smaller micelle diameter (Schmidt, 1980).

While the compositions and size of the casein micelle can be measured with various analytical techniques, the structure of casein micelles has not been fully elucidated and in agreement among scientists (Dalgleish and Corredig, 2012). The two popular theories demonstrating micellar structure of casein are submicelle and nanocluster models. The submicelle model, proposed by Schmidt in 1980, hypothesized that casein micelles consisted aggregates of casein protein linked by calcium phosphate. Internally, the submicelles are predominately formed by protein-protein interactions of caseins. On the other hand, the nanocluster model, proposed later by Holt et al. (1992), theocratized the micelle structure based on the casein phosphopeptide and calcium phosphate interactions, forming small and stable clusters. The enlargement of casein micelle relies on the crosslinking of nanoclusters by phosphorylated α_s-caseins and noncovalent bonds. Due to the lack of phosphate centers in κcaseins, they will remain on the surface of the micelle as observed from multiple studies (Dalgleish and Corredig, 2012).

Casein's role in cheese manufacture

The understanding of casein micelle structure provides foundation for altering physical, chemical, and microbiological environment of milk to produce various milk-derived food products such as yogurt, cheese. A representative casein-containing food is cheese. During cheese making, casein and fat are concentrated to convert milk into curd by acidification and rennet coagulation (Lucey and Kelly, 1994). Cheese making is a dehydration process as whey is drained from the curd to increase the total solid content in cheese. Followed by flavoring, molding, and aging, various types of cheese are produced for human consumptions (Fox, 1989). Figure 2-1 highlights casein's role in quantity and quality determination of cheese.



Casein in milk

Cheese manufacturing

Quantity determination

- Curd firmness
- Syneresis rate
- Moisture retention

Quality determination

- Flavor substrate
- Texture backbone

Figure 2-1 Overview of casein in milk determining quantity and quality of cheese.

Cheese yield determination

In the cheese industry, improving the quantity of cheese produced from a given amount of milk is a constant pursuit. Dairy food scientists have identified multiple factors that can influence the cheese yield. Milk compositions, namely the amount of protein and fat, are highlighted in multiple studies as indicators to cheese yield (Barbano and Sherbon, 1984; Emmons and Modler, 2010). Protein and fat are the two major solid components in cheese, so understanding their amounts in cheese milk can provide direct estimation of the final cheese solids. Protein-to-fat-ratio (PFR) has been developed as a reference metrics to understand cheese milk. A PFR range from 0.70 to 1.15 has resulted in differences in cheese composition and yield (Guinee et al., 2007). Since casein is the primary proteinaceous residual in cheese, knowing precisely the casein content in cheese milk can enhance the prediction of cheese yield. Multiple cheese yield predictive equations were published, and all of the models included compositions of fat and casein in cheese milk as factors (Emmons and Modler, 2010). Several later-identified cheese yield determinants included curd firmness, syneresis rate, and moisture retention. These qualities of curd or during cheese making have been linked back to the casein content in the cheese milk. Standardization of cheese milk became a common practice in cheese manufacturing. The standardization procedure serves to meet the legal definition of specific cheese varieties and to maximize the cheese yield without losing excessive fat and casein into whey (Lucey and Kelly, 1994). By adjusting the casein-to-fat ratio, the standardized milk can produce consistent and high yield for a given variety of cheese. Therefore, knowing the casein content in milk is a key to control cheese yield.

Cheese quality determination

On top of determining cheese yield, casein in milk has been linked to various quality determinations in cheese. Many types of cheese need to be ripened for their characteristic flavors and textures (Fox, 1989). During ripening, casein provides substrates for proteases and cultures in cheese to develop proteolysis-induced flavors. From casein hydrolysis, peptides, amino acids, acetic acid, ammonia, pyruvate, aldehydes, alcohols, carboxylic acids, and sulfur compounds were produced, contributing to the overall cheese flavor development (Singh et al., 2003). With the advancement of nontarget omics techniques, genomics and metabolomics studies have linked specific lactic acid bacteria to the desirable cheese flavor compounds (Steele et al., 2013). Proteolysis-induced flavor development is a key pathway for the overall flavor profile of a given cheese, and casein in cheese milk is the substrate for achieving the desirable cheese products.

Meanwhile, during cheese ripening, texture of cheese also changes due to proteolysis. Intact casein in cheese refers to the casein that has not been hydrolyzed. Intact casein content in the finished cheese has been correlated to cheese rheological and functional properties (Fenelon and Guinee, 2000). It provides structural backbone to cheese and contribute to the meltability and spreadability of cheese. To produce certain cheese texture, curd firmness and moisture retention should be controlled, and casein in cheese milk again determines these quality parameters. Intact casein's functionality extends to manufacturing processed cheese as researches have suggested that it can determine processed cheese's functional and textual properties (Kapoor and Metzger, 2008). Since hydrolyzed casein can produce flavors, while unhydrolyzed casein can provide structure, understanding the intact casein level in natural cheese can determine both flavor and textual profiles of the finished processed cheese.

Techniques to concentrate casein content in cheese milk

Membrane filtration aims to physically separate compounds based on their sizes. With the growing understanding of casein's role in cheese, membrane filtration techniques served as a complementary process to cheese manufacturing as it can alter casein content in cheese milk (Lipnizki, 2010). The early attempts of membrane filtration in cheese making involved ultrafiltration (UF). UF can physically separate milk protein from the other smaller compounds in milk such as lactose and minerals. The first application of UF in cheese making was done by Maubois et al. (1969), and the patented method concentrated milk protein by 5 to 7 times prior to cheese making. The method resulted in low whey draining and high-quality cheese curd. UF concentration of cheese milk can increase yield and improve cheese vat utilizations (Kumar et al., 2013). However, the quality of cheese produced by UF milk appeared some differences from the conventional cheese. The reduced mineral and retained undenatured whey protein created different flavor profiles during semi-hard and hard cheese ripening process (Lelievre and Lawrence, 1988). Because of product inconsistency in aged cheese, UF has been more successfully applied to fresh cheese production of Feta and fresh acid-curd varieties (Henning et al., 2006).

Another membrane processing technique known as microfiltration (MF) has also been applied to cheese manufacturing. Unlike UF, MF only retains micellar casein during the concentration process, which limits the whey protein retention issue found in UF-made cheese. Applications of MF-made cheese were published by Brandsma and Rizvi (2001), and the researchers manufactured Mozzarella cheese from 7-time concentrated MF milk. The study found that rheological properties of the MF-made Mozzarella cheese were improved compared to the control. The low whey protein retention in the MF-made Mozzarella also decrease the degree of proteolysis as compared to the control Mozzarella samples. In another study, Neocleous et al. (2002) used low concentration factor MF milk to make cheddar cheese. By standardizing the casein-to-fat ratio, the study found that MF can increase cheddar cheese yield, and the yield increase is independent from the concentration factor of MF. Various high-casein retentates from MF and UF increased cheese yield and maintained textural and flavor properties. Concentrating casein served as a backbone theory to the membrane processing applications in cheese manufacturing.

Measuring casein in milk

Due to the important role of casein in cheese manufacturing, determining casein content has been a focus of analytical method development. Quantification of casein in milk offers direct indication to fluid milk quality and cheese yield. The official method of casein measurement in milk relied on the nitrogen-titrimetry method, Kjeldahl. Lynch et al. (1998) published a collaborative study to standardize and evaluate the Association of Analytical Chemists (AOAC) method for casein content in milk. The method involved two approach to directly or indirectly determine casein content in milk. The principle of the method relied on isoelectrically precipitating casein in milk and separate casein from the non-casein fraction of milk by filtration. Casein content can either be quantified by directly measuring the isolated casein solids or indirectly calculating the difference between total nitrogen and non-casein nitrogen. This Kjeldahl-based method resulted in good repeatability and reproducibility, serving as the industry standard method for casein quantification since 1998.

Meanwhile, multiple chromatographic methods of casein analysis are available to separate and quantify casein using high performance liquid chromatography (HPLC). Separations based on reverse phase and size exclusion have been developed to fractionate β-casein, κ-casein, αs_1-casein, and αs_2-casein (Dimenna and Segall, 1981; van der Ven et al., 2001; Bonfatti et al., 2008). External calibrations were established with casein standards to determine casein content in milk. Similarly, capillary electrophoresis-based method was also developed to quantify serum protein and casein in milk (Recio and Olieman, 1996). While the primary goal of these methods was for protein separation, quantification of casein can also be achieved.

Automation and rapid sensing development in food analysis have produced several rapid quantification methods for casein. These methods are mostly spectroscopic methods involving, ultraviolet (UV) spectroscopy, infrared spectroscopy, and nuclear magnetic resonance (NMR) spectroscopy. Table 2-1 provides a summary of the rapid casein quantification methods using spectroscopies. The methods relied on calibration of known casein samples from either powder standards or milk. Near infrared spectroscopy (NIR) has been approved as a standard method for protein quantification in foods. Barbano and Dellavalle (1987) used the indirect approach to quantify total protein in milk and the non-casein protein using NIR. The calibration of NIR-based method was referenced to the Kjeldahl official method. The result of the study suggested that the mean casein content measured from the NIR-based method was not significantly different from the reference method ($P < 0.05$).

With the increasing consumer demand for ultra-high-temperature (UHT) milk, accurately quantifying casein in UHT milk created challenges for the precipitation-based methods because of casein and whey protein interactions (Belloque and Ramos, 2002). Belloque and Ramos developed a ^{31}P-NMR spectroscopy-based method to quantify casein in different pasteurized and UHT-processed milk. The method relied on external calibration of milk powder, and it required some sample preparations steps for the ^{31}P-NMR detection. The measurement range of this study was reported from 2.39 to 2.82%. Although there was no accuracy of the method reported, a comparison casein value obtained from the Kjeldahl official method was given.

Lüthi-Peng and Puhan (1999) published a method of using the 4th derivative of UV spectra to quantify milk protein and casein content in milk. The method involved a sample preparation step of unfolding milk proteins in guinidine-hydrochloric acid. Calibration was done to measure total protein and whey protein in milk, and casein content was then calculated. The method showed no significant effects to homogenization, preservation, or hydrolysis with protease, and protein and casein contents can be obtained from one UV spectroscopy scan. The accuracy of the method was compared to the Kjeldahl official method using linear model and the standard error of the method was found to be 0.06% casein based on a sample size of 34 with casein ranged from 1.92 to 3.34%.

Fourier-transform infrared (FTIR) spectroscopy were developed as a non-destructive method to measure various food components. Several applications have been published on casein quantifications. Hewavitharana (1997) first reported a FTIR-based method for casein quantification on raw milk samples. The method was developed using multivariate statistical models and validated with a set of 20 raw milk samples. The measurement range of the method was from 2.71 to 3.62% casein in raw milk, and the error of the method was reported from 0.08 to 0.1%. A follow-up study was done by Luginbühl (2002) using standard milk samples. In this study a larger casein measurement range was found, which also increased the measurement accuracy, lowering the error to 0.046-0.08%. It appeared that the increase of casein calibration range and high sample homogeneity resulted in accuracy improvement in FTIR-based measurements of casein. A mid infrared (MIR)-based method was published by McDermott et al. (2016) aiming for milk casein and free amino acid quantifications. The method was designed to capture casein content variation from different genetic breeds. Though the range of the casein measurement was not reported, the error from the study was 0.48%, almost 10 times higher than the FTIR method. Overall, the rapid methods focused on minimal sample preparation with various calibration ranges. In general, the increase of calibration range of casein improved the accuracy of the method.

Table 2-1 Summary of spectroscopic casein quantification methods

Principle	Sample preparation	Measurement range	Accuracy	Reference
NIR	None	NA	Mean difference remained insignificant ($P < 0.05$)	(Barbano and Dellavalle, 1987)
^{31}P-NMR	Methylenediphosphonic acid (internal standard) and EDTA	2.39 to 2.82%	NA	(Belloque and Ramos, 2002)
UV	Guinidine-hydrochloric acid	1.92 to 3.34%	SEP = 0.06%	(Lüthi-Peng and Puhan, 1999)
FTIR	None	1.8 to 4.5%	SEP = 0.046 to 0.08 %	(Luginbühl, 2002)
FTIR	None	2.71 to 3.62%	RMSE = 0.08 to 0.1%	(Hewavitharana, 1997)
MIR	None	NA	RMSE = 0.48%	(McDermott et al., 2016)

NIR = Near infrared, ^{31}P-NMR= Phosphorus-31 nuclear magnetic resonance, UV = ultra-violet, FTIR = Fourier-transform infrared, MIR = Mid-infrared, SEP = Standard error of prediction, RMSE = root mean square error.

Fluorescence spectroscopy

With the growing consumer demands on food quality in the recent years, the industrial food production continues to seek innovative analytical methods with high accuracy and efficiency (Karoui and Blecker, 2011). Compared to the traditional methods, spectroscopic instrumentations can increase detection speed, provide non-invasive analysis, and reduce labor and time while providing promising analytical results. In dairy productions, rapid detection techniques have been applied to compositional analysis of cheese, milk, milk powders. and milk curd synthesis using spectroscopic methods (Slobodan and Yukihiro, 2001; Fagan et al., 2011). Also, thanks to the nondestructive detection mode of the spectroscopic instrumentations, these methods can be used on-line or at-line to monitor quality parameters during production in real-time. Among the various spectroscopic methods developed by food researchers, fluorescence spectroscopy (FS) is known for its additional sensitivity compared to others. Studies have shown that FS is capable to detect food analytes in the sub-parts per billion range (Andersen and Mortensen, 2008).

Basic principles and instrumentations

The high sensitivity of FS comes from the measurement of the fluorescent molecules in foods known as fluorophores. The Jablonski diagram (Figure 2-2) illustrates the principle of FS. The fluorophore is excited by light absorption changing its energy state from ground state (S_0) to excited state (S_1). The excitation process resulted in a vibrational relaxation for the fluorophore, meaning that the molecule needs to transfer from the excited state to the lower energy state (with no radiation). When the molecule is in such process, the electron returns to the low energy state and emits light as the form of energy returned from the excitation. Based on Figure 2.2, the energy of emission is lower than the energy of excitation, meaning that the emission wavelength of a fluorophore is higher than its excitation wavelength. Such phenomenon is known as Stokes Shift. Therefore, fluorescence of a given molecule is described by both excitation and emission wavelength. The use of two wavelength parameters allows better specificity of the method, compared to other spectroscopic techniques which requires only one wavelength parameter. Also, because of the Stokes Shift, little to none spectral interference occurs to the measurement of emission wavelength, resulting in high resolution spectral collection, which translates to high sensitivity of the analyte measurement (Ahmad et al., 2017). However, the major limitation of the FS is that the analyte of FS must be a fluorophore due to unique physical principle. Certain analytes may not carry fluorescent properties, so they are not suitable for FS analysis.

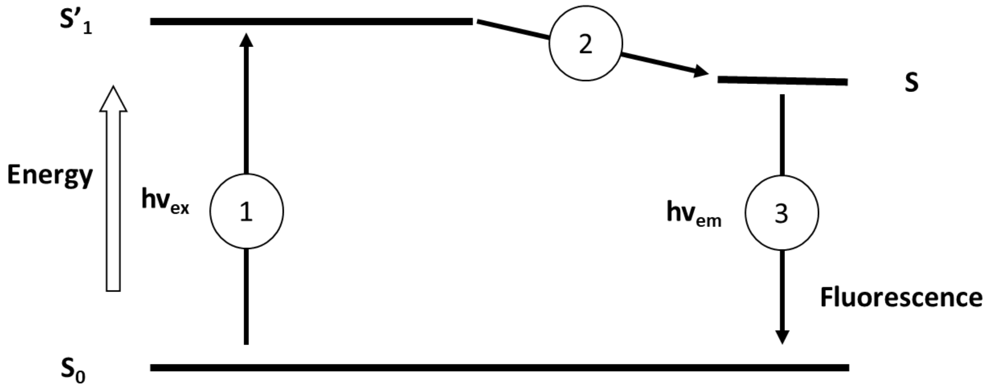


Figure 2-2 Jablonski diagram showing the basic principle in fluorescence spectroscopy (adapted from Karoui and Blecker, 2011).

To obtain fluorescence spectra, a dual monochromator setup has been widely used. Figure 2-3 provides a schematic demonstration of FS measurement process. In short, fluorophores are excited by a monochromator at certain wavelengths, and the fluorophore-containing sample emits light at a different wavelength, which can be filtered and detected by the photodetector. Spectral data are then collected into computers and multivariate statistical models can be constructed based on various chemometric analyses. There are two orientations of the FS sample holder. The right-angle position requires that the excitation light travels through the sample from one side, and the detector is positioned at right angles to collect the emission signal. The right-angle orientation provides sensitive measurement of fluorophore at low concentration. With an emission absorbance less than 0.1, the analyte concentration is proportional to the emitting light intensity (Karoui and Blecker, 2011). However, in food analysis, since the matrix contains multiple components with various levels of fluorophore's presence, right-angle measurement is limited by scattering and high emission absorbance. Therefore, an alternative orientation known as the front-face fluorescence spectroscopy (FFFS) is favored due to its ability to measure turbid samples. FFFS compromises some analytical sensitivities as its emission spectra are often nosier than those obtained from right-angle orientations. In food analysis, FFFS is used to capture changes in food processing and rapidly estimate compositions, so reasonable loss of sensitivity may not affect its intended uses.

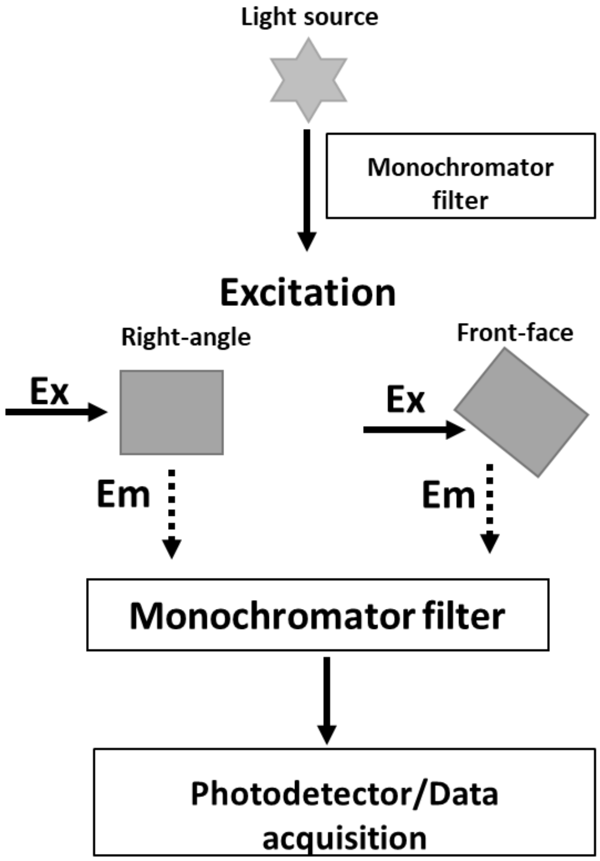


Figure 2-3 Basic setup of a fluorescence spectroscopy (adopted Karoui and Blecker, 2011).

Fluorescence spectra

Since there are two monochromators involved in the measurement of FS, various types of spectra can be obtained from the same instrument. The fundamental type of spectrum obtained from FS measurement is described in the Jablonski diagram (Figure 2-2). With a fixed excitation wavelength, the emission monochromator can scan a range of wavelength to capture a distribution of light energy emitted from a single excitation. This type of spectrum is known as the emission spectrum. Using an opposite approach, by fixing the emission monochromator at a single wavelength and varying the excitation wavelength, a fluorophore will be excited with different wavelengths and only emit light in response to some of them. This type of FS spectrum is known as the excitation spectrum. Both the emission and excitation spectra are commonly used to measure known fluorophores (Karoui and Blecker, 2011). For example, in the early studies of FFFS characterization of dairy foods, Herbert et al. (1999) used the known FS parameters with excitation of 290 nm and emission wavelength ranged from 305 to 450 nm to measure coagulation of milk. The FS parameters were referenced to the pure tryptophan fluorescent properties without product-specific fluorescence identification.

With the development of FS instrumentations, a non-targeted approach of FS measurement was more commonly used known as excitation-emission matrix (EEM). EEM sometimes is referred as the 3-deminsional or 3-way measurement of fluorescence. It is produced by collating continuous scans of emission spectra from a range of excitation wavelengths (Bahram et al., 2006). EEM contains large amount of data highlighting all available fluorophores in a sample. Therefore, it is normally used as an exploratory tool to understand the specific food sample and select the most useful fluorophore as the marker compound. Kokawa et al. (2015) used EEM to understand the distribution of intact casein in cheese, and the authors identified the tryptophan difference in aged and non-aged cheese, which was used to predict the intact casein concentrations. Sometimes, there are more than one fluorophore to be included in the analysis, and to rapidly collect fluorescence information from all fluorophores, synchronous fluorescence is developed. Synchronous fluorescence spectra are obtained by plotting fluorescence intensity against the excitation/emission wavelength combinations. In this way, spectra selectivity is increased especially dealing with multi-fluorophore samples. For example, fluorescence spectral information from riboflavin, oxidized fatty acids, and vitamin E was used to authenticate virgin olive oil (Poulli et al., 2007). The synchronous fluorescence spectra provided reliable prediction and avoid potential scattering effects in normal FS measurement.

Fluorophores in foods

There are multiple intrinsic fluorophores in foods for FS analysis. These fluorophores are naturally present in food products or have been synthesized during food processing. Representative fluorophores and their excitation/emission wavelengths are summarized in Figure 2-4. Since food is a dynamic chemical environment, measuring fluorophore concentrations in foods can provide both analysis of food compositions and characterization of processing and storage changes. Specifically, the source of foods determines the fluorophores present and the information generated from FS measurements. In Figure 2.4, fat-containing foods such as whole milk and meat carry riboflavin and vitamin A, and researchers have developed FS-based methods to measure these fluorophores and correlate with other quality parameters. For example, the change in riboflavin in whole milk serves as an indicator to light exposure, so an FS method measuring riboflavin in whole milk provides estimation for light exposure of raw milk (Choe et al., 2005). Another characteristic fluorophore in foods is chlorophyll. Naturally present in plants, chlorophyll has been used to characterize apple fruit maturity and quality (Noh and Lu, 2007; Cerovic et al., 2008). Due to the unique excitation/emission wavelength combination of chlorophyll, minimal spectral interference occurs to its measurement. Therefore, trace amount of chlorophyll can provide reliable fluorescence signals. Chlorophyll is then used as a marker to discriminate between grass-fed milk and conventional milk using FS techniques (Bhattacharjee et al., 2018).

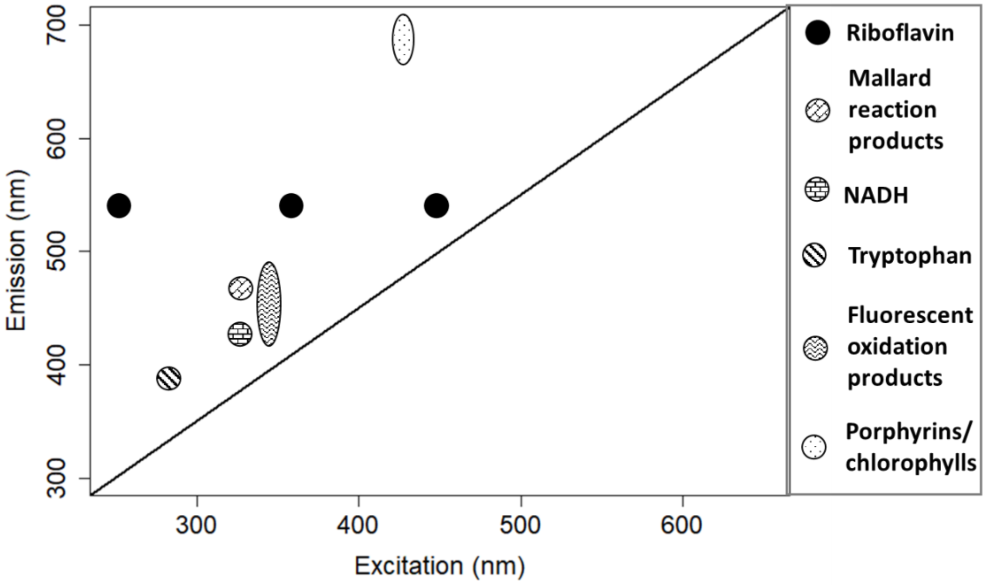


Figure 2-4 Excitation and emission maxima of fluorophores present in food products (adapted from Andersen and Mortensen, 2008).

Also, some fluorophores may not be naturally present in fresh foods, but they can be formed during food processing and storage. Fluorescent oxidation products and Maillard browning products are examples of such molecules. Oxidation of foods occurs when certain food compounds are exposed to oxygen, and such process may result in undesirable quality changes of foods. FS have been used as a tool to monitor such changes by using fluorescent oxidation products as markers. FS measurement of these products occurs in a wider emission wavelength range from 400 to 500 nm, and these compounds are derivatives of riboflavin oxidation and fatty acid oxidation primarily found in meat and oil products (Guimet et al., 2005; Gatellier et al., 2007). Similarly, during heat treatment of foods, Maillard reaction occurs and results in browning products. These products contain cyclic structures, giving them fluorescent properties. Researchers have been using the Maillard products as markers to characterize heat treatment and storage of food products. One of the representative fluorescence-based parameters is known as the FAST (Fluorescence of Advanced Maillard products and Soluble Tryptophan) index developed by Birlouez-Aragon et al. (2002). The FAST index calculates the ratio between the advanced Maillard products and soluble tryptophan in dairy products and estimate the heat load of them.

Table 2-2 Example studies of tryptophan fluorescence for dairy food quantification and processing-induced changes characterizations

Product	Function	Measurement	Spectral parameters	Reference
Milk	Quantification measurement	β-lg and alkaline phosphatase	Ex at 290 nm and em ranged 320-360 nm	(Kulmyrzaev et al., 2005)
Cheese		Intact casein	Ex at 290 nm maximum em at 345 nm	(Kokawa et al., 2015)
Skim milk powder		Lactulose	Ex at 290 nm, em at 307 and 324 nm	(Ayala et al., 2017)
Milk	Processing-induced changes	Estimation of heat treatment	Ex at 290 nm, em at 340 nm	(Kulmyrzaev et al., 2005)
Non-fat dry milk		Storage temperature effects	Ex at 290 nm, and em ranged 305-450 nm	(Liu and Metzger, 2007)
Cream cheese		Formulation variations	Ex ranged from 260-360 nm, em ranged 280-600 nm	(Andersen et al., 2010)
Emmental cheese		Cheese ages and origins	Ex at 290 nm, em ranged 305-400 nm	(Karoui et al., 2006)

Ex = excitation wavelength; Em = emission wavelength.

A more abundant fluorophore in foods is tryptophan in peptides and proteins. Tryptophan is therefore extensively studied in protein-containing foods for quantification measurement and processing-induced quality changes. Table 2-2 provides some examples of tryptophan-based FS measurements of dairy foods. For protein detections, Kulmyrzaev et al. (2005) developed characterized heat treatments of milk and its protein denaturation using FFFS. The study collected tryptophan and NADH fluorescence spectra and detect protein changes based on heating time and temperature of the milk. The models were able to predict alkaline phosphatase and β-lactoglobulin levels using the FFFS spectra. Though in different food matrix, the spectral parameters for these studies were similar, aimed to measure tryptophan. The slight variation in tryptophan measurement parameters provides specificity to each method, and researchers have used multiple methods to select the best-performing wavelengths.

Chemometrics overview

Since most of the FFFS measurement of food samples results in complicated and noisy spectra, statistical models are required as a part of the method development process to ensure the accuracy and reproducibility. Multivariate statistical analyses (also known as chemometrics) are needed to extract quantitative, qualitative, or structural information from these spectra (Karoui et al.,2010). It provides variable reductions and calibrations of the spectral data to the reference data, enhancing values of the fluorescence measurements. In general, chemometrics methods can be divided into unsupervised and supervised analysis. Unsupervised methods focus primarily on pattern recognition of the spectral data. Because fluorescence spectra of foods contain multiple components, unsupervised methods are helpful to identify individual component’s contribution to the obtained spectra. On the other hand, supervised methods aim to manipulate spectral information and predict certain reference attributes. It involves multivariate model development and validation to connect spectral information to other chemical and physical properties of foods. Both approaches provide invaluable knowledge and applications to food analysis, and they establish the foundation for spectroscopic analysis of foods.

Unsupervised analysis

Unsupervised chemometric tools focus to decode the complex spectra obtained from measuring food samples. Principal component analysis (PCA) has been one of the most common methods of compressing and visualizing 2-dimentional spectral data (Karoui et al., 2010). In the data matrix, fluorescence intensity of each corresponding wavelength is transformed into latent variables by maximizing the covariance among the wavelength. PCA reduces the number of variables in the original data matrix with minimal loss of the spectral information. The principal components (PC), can be associated with specific chemical and structural components in the food matrix such as analyte concentrations and protein-protein interactions (Herbert et al., 1999; Kulmyrzaev et al., 2005). PCA also provides data visualizations for high-dimensional data. Because the PC explain majority of the spectral variance, plotting the first and second PC can often visualize sample differences among treatments. This visualization tool has been used to classify samples with different processing process, geographic origins, and formulation differences (Table 2.2).

While PCA has been the dominating unsupervised method in FFFS studies, several other unsupervised methods have been applied to analyze fluorescence spectra. For example, hierarchical clustering analysis (HCA) is a clustering method based on linkage and distance calculations. Various linkage and distance methods can reduce high dimensionalities and visualize sample similarities on a dendrogram. Several studies have applied HCA to find linkages between edible and lampante virgin olive oil and between brandies and wine distillate (Poulli et al., 2005; Sádecká et al., 2009). The flexibility in distance and linkage calculations empower HCA as an efficient pattern recognizing tool for unknown samples.

In the meantime, because of the special data structure of the EEM generated by FFFS, researchers have applied multi-way analysis to directly analyze the 3-dimensional data. Parallel factor analysis (PARAFAC) is a well-known method developed by Bro (1999). The complicated EEM data array is decomposed into several 2-way spectrum known as parallel factors.

PARAFAC is considered as a unsupervised method because the parallel factors produced by the calculation has been well-correlated to the individual fluorophores in the sample mix (Bro, 1999). Therefore, for exploratory purposes, the non-targeted EEM approach combined with PARAFAC can collect total fluorescence in a sample and decompose the specific fluorophores in it. Using this approach, researchers have studied specific fluorophores in yogurt and honey, which generated promising classification power (Christensen et al., 2005; Lenhardt et al., 2015). However, collecting EEM is relatively time-consuming compared to excitation/emission scans, so for relatively simple food matrix, EEM coupled with PARAFAC may not be efficient for method development. Also, the number of parallel factors needs to be specified prior to running the analysis, meaning that researchers should have some general knowledge on the sample rather than blindly relying on the algorithms.

Supervised modelling

Unsupervised tests offer exploratory analysis of spectral data, and they can provide extensive understanding of the samples. To expand the use of spectral data, supervised methods can establish predictive models between a sample's spectrum and its reference values. Overall, supervised methods can be categorized into classification and regression. The two types of supervised methods rely on different reference values. For classifications, the reference values are categorical, and the predictive models are designed to classify unknown samples into the trained classes. On the other hand, regression models are developed on continuous data, and the models are designed to predict specific values from the spectral input. Though with different goals, supervised chemometric modelling has a general workflow shown in Figure 2-5.

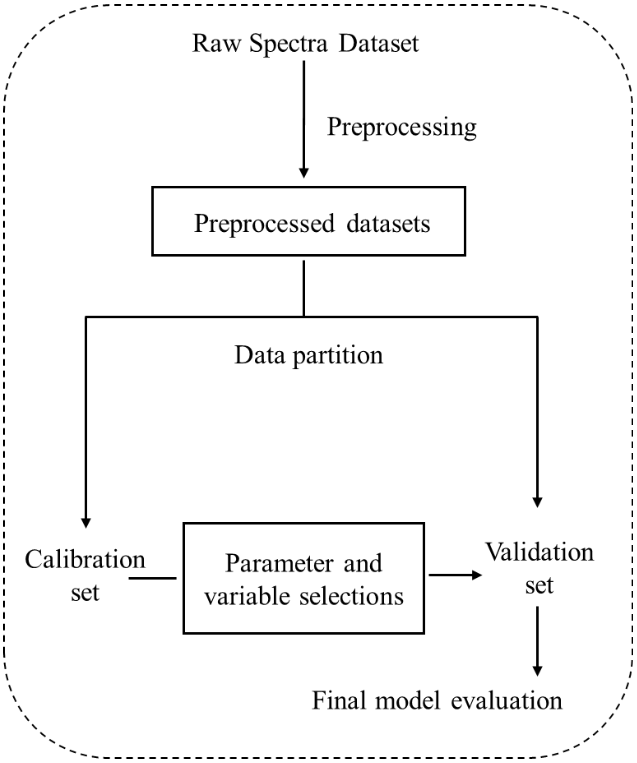


Figure 2-5 General workflow of supervised chemometrics modelling.

The raw spectral data are normally preprocessed to reveal spectra overlapping and instrumental noise (Brown et al., 2000). Numerous preprocessing tools are developed by spectroscopists over the years. In general, preprocessing techniques have several categories: normalization, derivation, smoothing, and corrections. Although researchers claim to use the same spectroscopic technique (e.g. FFFS), the sample preparation, instrumental setup, and ambient environment can create much variation in the raw spectra collected. Therefore, the fundamental goal of preprocessing is to unify spectra and minimize external effects to the spectral variation. There is no “golden rule” of using the best preprocessing technique, rather a trial-and-error approach is recommended by many researchers (Coronel-Reyes et al., 2018). In FFFS studies, preprocessing techniques have not been applied as widely as they are in infrared spectroscopies (IR). One possible reason for it may be that FFFS collects spectral information from the fluorophores, which yields more specific spectral pattern than IR.

After preprocessing, the spectra datasets are partitioned into calibration and validation sets. Though there is no hard rule on the partitioning proportion, a common practice is to randomly select 70-80% of the samples in the calibration set and leave 20-30% in the validation set. The calibration set is used to develop and tune the predictive models, and the validation set is a group of unknown samples intended to test the robustness of the developed model (Wold et al., 2001). Multiple models can be developed using the calibration set, and model performance is evaluated by its error to the reference data via cross validations (CV). CV are internal tests to the model, and the results from CV should be validated externally to ensure its robustness (Geladi, 2002). The evaluation metrics such as root mean square error of cross validation (RMSECV) and accuracy are used to tune the model parameters. For example, RMSECV often determines the number of latent variables for partial least square regression (PLSR). The optimal number of hidden layers can be chosen from the highest classification accuracy generated by artificial neutral network (ANN) classifiers. Some studies did not include the validation step in the model development and designed the experiments as a “proof-of-concept”. However, it is less likely to observe recent publications without the validation step in the field of rapid method development.

Additionally, to further improve the model performance, variable selections are applied to the preprocessed spectra. The aim of variable selection is to determine a subset of the variable (wavelength) which can minimize the prediction error compared to the model developed with the full spectra (Mehmood et al., 2012). For example, to develop PLSR models, several novel approaches of variable selections have been published including backward elimination, genetic algorithm, interval PLSR, and elastic net regression (Wang et al., 2018). While variable selection have been less commonly found in FFFS studies, Teófilo et al. (2009) demonstrated that ranking variables using regression vector, correlation vector, and variable influence on projection can enhance PLSR model performance of a sample fluorescence dataset. Similar approaches have been found to determine butter adulteration using the successive projection algorithm (Dankowska et al., 2014). In this study, even though the author did not find prediction improvement using variable selection, the selection algorithm identified the most contributing variables, which revealed the determining emission wavelengths for confirming the butter adulteration. Therefore, variable selection techniques can also identify the most informative wavelength from the dataset and connect statistical models to the chemical founding principles in foods.

Predictive FFFS studies

In the validation step, established models with tuned parameters and selected variables are evaluated with unknown samples. Table 2-3 provided some recent publications of FFFS prediction studies with external validations. The quantifications and classifications focused on authentication and quality parameters, and a common objective for all these studies is to provide rapid analysis for the food industry.

Quantifying adulterants and quality parameters in foods has been a focus of recent predictive FFFS publications. Maintaining good reproducibility of FFFS-based method is a challenge as the chemometric models are prone to overfitting in the calibration step (Wold et al., 2001). Therefore, good validation steps ensure functional measurement and spectral processing that can be generalized. Tan et al. (2017) developed an authentication FFFS-based method including quantifying the percentage of refined frying oil in pure vegetable oil using PLSR model. The prediction error was evaluated first using RMSECV during the calibration step, and the validation step yielded a similar RMSEP to RMSECV. The authors then concluded the method’s validity as a rapid authentication tool for vegetable oil. In a pure statistical sense, RMSEP should be the best way of accessing prediction error (Fearn, 2002). Sometimes, secondary parameters are derived based on the RMSEP to enhance understanding of the error with the specific dataset. These parameters include coefficient of determination (R^2), residual prediction deviation (RPD), sensitivity, and specificity. Tan et al. (2017) used R^2 to facilitate audience's understanding of the model performance as more people are familiar with R^2 than RMSEP. An almost identical approach was taken by Markechová et al. (2014) to authenticate brandy. The R^2 included in both studies was above 0.85, indicating useful model predictions in food analysis. However, as some researchers have pointed out, R^2 describes linearity between the predicted and reference values in a model, which does not necessarily equal to error. Also, R^2 is sensitive to range of the validation set, making it difficult to judge prediction performance of sample sets with low standard deviations (Twomey, 2006).

Oto et al. (2013) developed a FFFS-based method to quantify ATP content and plate count on pork meat surface. The quality parameters prediction is another popular approach of using FFFS spectra for prediction purposes. The study measured several fluorophores on pork meat surface. Even though the spectra were not directly correlated to the measured compound, the FFFS spectra made responses based on the different ATP content and plate count in the calibration sample set. The study applied three different linear regression methods to predict the two quality parameters. The models yielded similar results and were externally validated for their reproducibility. Babu and Amamacharla (2018) conducted a study using a similar approach for MPC's solubility. Even though solubility is an empirical measurement of powder quality, it is still determined by fundamental dairy chemistry which involves changes in protein and Maillard browning reactions. The authors then made FFFS measurement to capture the changes in powders and developed a prediction model based on PLSR. The study included a secondary evaluation parameter called RPD, which was first mentioned by Williams and Norris (2001). This parameter is the ratio between the standard deviation of the validation set and the RMSEP, which describes the overall model prediction power.

FFFS-based methods can also provide classifications in foods and beverages. FFFS spectra were used to develop classification models for sherry vinegar and Argentine white wines (Callejón et al., 2012; Azcarate et al., 2015). For classification studies, accuracy is the primary evaluation parameter for the models, and accuracy (correct classified percentage) for the validation set provides confirmation of the model's robustness. In addition, specificity and sensitivities can be used to understand the portion of false positives and true negatives in the model. There are many classification models that have been used in food science studies. They include partial least square discriminant analysis (PLS-DA), support vector machine (SVM), Soft independent modelling of class analogies (SIMCA), and linear discriminant analysis (LDA).

Overall, FFFS combined with chemometrics has shown to be an effective tool for rapid food analysis. In this thesis, two applications of FFFS were proposed to measure casein in fluid milk and detect protein leak in UF permeate stream. Chapter 3 introduces the specific objectives of this thesis. A preliminary study in chapter 4 developed and validated a FFFS-based casein quantification method in raw milk. Chapter 5 extended the preliminary method and developed a method to measure both casein and casein-to-crude-protein ratio using FFFS and chemometrics. In chapter 6, FFFS and chemometrics were applied to understand protein leak during UF processing and developed predictive models to determine true protein content in UF permeate and identify the source of protein leak.

Table 2-3 Examples of recent FFFS prediction studies in food science involving external validation

Product	Vegetable oil	Brandy	Pork meat	Milk protein concentrate	Sherry vinegar	Argentine white wines
Objective	Quantification of refined frying oil as an adulterant in vegetable oil	Quantification of mixed wine spirit in brandy as an adulterant	ATP content and plate count quantifications	Quantification of solubility	Classification based on sherry vinegar categories	Classification based on grape varieties
Model	PLSR	PLSR	PLSR, MLR, PCR	PLSR	PLS-DA, SVM	SIMCA, PLS-DA, SPA-LDA
Model evaluation parameters	RMSECV, RMSEP, R^2	RMSECV, RMSEP, R^2	RMSECV, RMSEP	RMSECV, RMSEP, RPD, R^2	Accuracy, sensitivity, specificity	Accuracy, sensitivity, specificity
Reference	(Tan et al., 2017)	(Markechová et al., 2014)	(Oto et al., 2013)	(Babu and Amamcharla, 2018)	(Callejón et al., 2012)	(Azcarate et al., 2015)

PLSR = partial least square regression, MLR = multiple linear regression, PCR = principal component regression, PLS-DA = partial least square discriminant analysis, SVM = support vector machine, SPA-LDA = successive projection algorithm – linear discriminant analysis, RMSECV = root mean square error of crossvalidation, RMSEP = root mean square error of prediction, R2 = coefficient of determination, RPD = residual prediction deviation.

References

Ahmad, M.H., A. Sahar, and B. Hitzmann. 2017. Fluorescence spectroscopy for the monitoring of food processes. Adv.Biochem. Eng./Biot. Springer, Cham.

Andersen, C.M., and G. Mortensen. 2008. Fluorescence Spectroscopy: A Rapid Tool for Analyzing Dairy Products. J. Ag. Food. Chem. 56:720-729.

Andersen, C.M., M.B. Frøst, and N. Viereck. 2010. Spectroscopic characterization of low- and non-fat cream cheeses. Int. Dairy J. 20:32–39.

Ayala, N., A. Zamora, C. González, J. Saldo, and M. Castillo. 2017. Predicting lactulose concentration in heat-treated reconstituted skim milk powder using front-face fluorescence. Food Control. 73:110–116.

Azcarate, S.M., A. de Araújo Gomes, M.R. Alcaraz, M.C. Ugulino de Araújo, J.M. Camiña, and H.C. Goicoechea. 2015. Modeling excitation–emission fluorescence matrices with pattern recognition algorithms for classification of Argentine white wines according grape variety. Food Chem. 184:214–219.

Babu, K.S., and J.K. Amamcharla. 2018. Application of front-face fluorescence spectroscopy as a tool for monitoring changes in milk protein concentrate powders during storage. J. Dairy Sci. 101:10844–10859.

Bahram, M., R. Bro, C. Stedmon, and A. Afkhami. 2006. Handling of Rayleigh and Raman scatter for PARAFAC modeling of fluorescence data using interpolation. J. Chemometr. 20:99–105.

Barbano, D.M., and J.W. Sherbon. 1984. Cheddar cheese yields in New York. J. Dairy Sci. 67:1873–1883.

Barbano, D.M., and M.E. Dellavalle. 1987. Rapid method for determination of milk casein content by infrared analysis. J. Dairy Sci. 70:1524–1528.

Belloque, J., and M. Ramos. 2002. Determination of the casein content in bovine milk by 31P-NMR. J. of Dairy Res. 69:411–418.

Bhattacharjee, U., D. Jarashow, T.A. Casey, J.W. Petrich, and M.A. Rasmussen. 2018. Using fluorescence spectroscopy to identify milk from grass-fed dairy cows and to monitor its photodegradation. J. Agric. Food Chem. 66:2168–2173.

Birlouez-Aragon, I., P. Sabat, and N. Gouti. 2002. A new method for discriminating milk heat treatment. Int. Dairy J. 12:59–67.

Bonfatti, V., L. Grigoletto, A. Cecchinato, L. Gallo, and P. Carnier. 2008. Validation of a new reversed-phase high-performance liquid chromatography method for separation and quantification of bovine milk protein genetic variants. J. Chromatogr. A 1195:101–106.

Brandsma, R.L., and S.S.H. Rizvi. 2001. Effect of manufacturing treatments on the rheological character of Mozzarella cheese made from microfiltration retentate depleted of whey proteins1. Int. J. Food Sci. Technol. 36:601–610.

Bro, R. 1999. Exploratory study of sugar production using fluorescence spectroscopy and multi-way analysis. Chemom. Intell. Lab. Syst. 46:133–147.

Brown, C.D., L. Vega-Montoto, and P.D. Wentzell. 2000. Derivative preprocessing and optimal corrections for baseline drift in multivariate calibration. Appl. Spectrosc. 54:1055–1068.

Callejón, R.M., J.M. Amigo, E. Pairo, S. Garmón, J.A. Ocaña, and M.L. Morales. 2012. Classification of Sherry vinegars by combining multidimensional fluorescence, PARAFAC and different classification approaches. Talanta 88:456–462.

Cerovic, Z.G., N. Moise, G. Agati, G. Latouche, N. Ben Ghozlen, and S. Meyer. 2008. New portable optical sensors for the assessment of winegrape phenolic maturity based on berry fluorescence. J. Food Comp. Anal. 21:650–654.

CFR - Code of Federal Regulations Title 21. Accessed March 11, 2017. https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?fr=131.110.

Choe, E., R. Huang, and D.B. Min. 2005. Chemical reactions and stability of riboflavin in foods. J. Food Sci. 70: R28–R36.

Christensen, J., E.M. Becker, and C.S. Frederiksen. 2005. Fluorescence spectroscopy and PARAFAC in the analysis of yogurt. Chemom. Intell. Lab. Syst. 75:201–208.

Coronel-Reyes, J., I. Ramirez-Morales, E. Fernandez-Blanco, D. Rivero, and A. Pazos. 2018. Determination of egg storage time at room temperature using a low-cost NIR spectrometer and machine learning techniques. Comput. Electron. Agr. 145:1–10.

Dalgleish, D.G., and M. Corredig. 2012. The structure of the casein micelle of milk and its changes during processing. Annu Rev Food Sci Technol. 3:449–467.

Dankowska, A., M. Małecka, and W. Kowalewski. 2014. Application of synchronous fluorescence spectroscopy with multivariate data analysis for determination of butter adulteration. Int. J. Food Sci. Technol. 49:2628–2634.

Dimenna, G.P., and H.J. Segall. 1981. High-performance gel-permeation chromatography of bovine skim milk proteins. J. Liqui. Chromatogr. 4:639–649.

Emmons, D.B., and H.W. Modler. 2010. Invited review: A commentary on predictive cheese yield formulas1. J. Dairy Sci. 93:5517–5537.

Fagan, C.C., T.G. Ferreira, F.A. Payne, C.P. O'Donnell, D.J. O'Callaghan, and M. Castillo. 2011. Preliminary evaluation of endogenous milk fluorophores as tracer molecules for curd syneresis. J. Dairy Sci. 94:5350–5358.

Fearn, T .2002. Assessing Calibrations: SEP, RPD, RER and R2. NIR News, 13: 12–13.

Fenelon, M. A., & Guinee T. P. 2000. Primary proteolysis and textural changes during ripening in Cheddar cheeses manufactured to different fat contents. Int. Dairy J. 10:151-158.

Fox, P.F. 1989. Proteolysis during cheese manufacture and ripening1. J. Dairy Sci. 72:1379–1400.

Gatellier, P., S. Gomez, V. Gigaud, C. Berri, E.L. Bihan-Duval, and V. Santé-Lhoutellier. 2007. Use of a fluorescence front face technique for measurement of lipid oxidation during refrigerated storage of chicken meat. Meat Sci. 76:543–547.

Geladi, P. 2002. Some recent trends in the calibration literature. Chemom. Intell. Lab. Syst. 60:211–224.

Guimet, F., J. Ferré, and R. Boqué. 2005. Rapid detection of olive–pomace oil adulteration in extra virgin olive oils from the protected denomination of origin "Siurana" using excitation–emission fluorescence spectroscopy and three-way methods of analysis. Anal. Chim. Acta 544:143–152.

Guinee, T.P., E.O. Mulholland, J. Kelly, and D.J.O. Callaghan. 2007. Effect of protein-to-fat ratio of milk on the composition, manufacturing efficiency, and yield of cheddar cheese. J. Dairy Sci. 90:110–123.

Henning, D.R., R.J. Baer, A.N. Hassan, and R. Dave. 2006. Major advances in concentrated and dry milk products, cheese, and milk fat-based spreads1. J. Dairy Sci. 89:1179–1188.

Herbert, S., A. Riaublanc, B. Bouchet, D.J. Gallant, and E. Dufour. 1999. Fluorescence spectroscopy investigation of acid-or rennet-induced coagulation of milk. J. Dairy Sci. 82:2056–2062.

Hewavitharana, A.K. 1997. Fourier transform infrared spectrometric method for the rapid determination of casein in raw milk. Analyst. 122:4.

Holt, C. 1992. Structure and stability of bovine casein micelles. In Advances in protein chemistry 43: 63-151. Academic Press.

Horne, D.S. 2006. Casein micelle structure: Models and muddles. Curr. Opin. Colloid Interface Sci 11:148–153.

Kapoor, R., & Metzger L. E. (2008). Process cheese: Scientific and technological aspects—A review. Compr. Rev. Food Sci. F. 7, 194-214.

Karoui, R., A.M. Mouazen, É. Dufour, L. Pillonel, E. Schaller, D. Picque, J.D. 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. 223:44–50.

Karoui, R., and C. Blecker. 2011. Fluorescence spectroscopy measurement for quality assessment of food systems—a review. Food Bioprocess Tech. 4:364–386.

Karoui, R., G. Downey, and C. Blecker. 2010. Mid-infrared spectroscopy coupled with chemometrics: a tool for the analysis of intact food systems and the exploration of their molecular structure−quality relationships − a review. Chem. Rev. 110:6144–6168.

Kokawa, M., S. Ikegami, A. Chiba, H. Koishihara, V. Trivittayasil, M. Tsuta, K. Fujita, and J. Sugiyama. 2015. Measuring cheese maturation with the fluorescence fingerprint. Food Sci. Technol. Res. 21:549–555.

Kulmyrzaev, A.A., D. Levieux, and É. Dufour. 2005. Front-face fluorescence spectroscopy allows the characterization of mild heat treatments applied to milk. Relations with the denaturation of milk proteins. J. Agric. Food Chem. 53:502–507.

Kumar, P., N. Sharma, R. Ranjan, S. Kumar, Z.F. Bhat, and D.K. Jeong. 2013. Perspective of membrane technology in dairy industry: a review. Asian-Australas. J. Anim. Sci. 26:1347–1358.

Lelievre, J., and R.C. Lawrence. 1988. Manufacture of cheese from milk concentrated by ultrafiltration. J. Dairy Res. 55:465–478.

Lenhardt, L., R. Bro, I. Zeković, T. Dramićanin, and M.D. Dramićanin. 2015. Fluorescence spectroscopy coupled with PARAFAC and PLS DA for characterization and classification of honey. Food Chem. 175:284–291.

Lipnizki, F. 2010. Cross-Flow Membrane Applications in the Food Industry. John Wiley & Sons, Ltd.

Liu, X., and L.E. Metzger. 2007. Application of fluorescence spectroscopy for monitoring changes in nonfat dry milk during storage. J. Dairy Sci. 90:24–37.

Lucey, J., and J. Kelly. 1994. Cheese yield. Int. J. Dairy Technol. 47:1–14.

Luginbühl, W. 2002. Evaluation of designed calibration samples for casein calibration in fourier transform infrared analysis of milk. LWT - Food Sci. Technol. 35:554–558.

Lüthi-Peng, Q., and Z. Puhan. 1999. Determination of protein and casein in milk by fourth derivative UV spectrophotometry. Anal. Chim. Acta. 393:227–234.

Lynch, J.M., D.M. Barbano, and J.R. Fleming. 1998. Indirect and direct determination of the casein content of milk by Kjeldahl nitrogen analysis: collaborative study. J AOAC Int. 81:763–774.

Markechová, D., P. Májek, and J. Sádecká. 2014. Fluorescence spectroscopy and multivariate methods for the determination of brandy adulteration with mixed wine spirit. Food Chem. 159:193–199.

Maubois, J. L., G. Mocquot, and L. Vassal. 1969. A method for processing milk and dairy products. French Pat. n° 2052.121.

McDermott, A., G. Visentin, M. De Marchi, D.P. Berry, M.A. Fenelon, P.M. O'Connor, O.A. Kenny, and S. McParland. 2016. Prediction of individual milk proteins including free amino acids in bovine milk using mid-infrared spectroscopy and their correlations with milk processing characteristics. J. Dairy Sci. 99:3171–3182.

Mehmood, T., K.H. Liland, L. Snipen, and S. Sæbø. 2012. A review of variable selection methods in Partial Least Squares Regression. Chemom. Intell. Lab. Syst. 118:62–69.

Neocleous, M., D.M. Barbano, and M.A. Rudan. 2002. Impact 9 of low concentration factor microfiltration on milk component recovery and cheddar cheese yield1. J. Dairy Sci. 85:2415–2424.

Noh, H.K., and R. Lu. 2007. Hyperspectral laser-induced fluorescence imaging for assessing apple fruit quality. Postharvest Biol. Technol. 43:193–201.

Oto, N., S. Oshita, Y. Makino, Y. Kawagoe, J. Sugiyama, and M. Yoshimura. 2013. Non-destructive evaluation of ATP content and plate count on pork meat surface by fluorescence spectroscopy. Meat Sci. 93:579–585.

Poulli, K.I., G.A. Mousdis, and C.A. Georgiou. 2005. Classification of edible and lampante virgin olive oil based on synchronous fluorescence and total luminescence spectroscopy. Anal. Chim.Acta. 542:151–156.

Poulli, K.I., G.A. Mousdis, and C.A. Georgiou. 2007. Rapid synchronous fluorescence method for virgin olive oil adulteration assessment. Food Chem. 105:369–375.

Recio, I., and C. Olieman. 1996. Determination of denatured serum proteins in the casein fraction of heat-treated milk by capillary zone electrophoresis. Electrophoresis. 17:1228–1233.

Sádecká, J., J. Tóthová, and P. Májek. 2009. Classification of brandies and wine distillates using front face fluorescence spectroscopy. Food Chem. 117:491–498.

Šašić, S., and Y. Ozaki. 2001. Short-wave near-infrared spectroscopy of biological fluids. 1. Quantitative analysis of fat, protein, and lactose in raw milk by partial least-squares regression and band assignment. Anal. Chem. 73:64–71.

Schmidt, D.G. 1980. Colloidal aspects of casein. Neth. Milk Dairy J. 34:42–64.

Singh, T.K., M.A. Drake, and K.R. Cadwallader. 2003. Flavor of cheddar cheese: a chemical and sensory perspective. Compr. Rev. Food Sci. F. 2:166–189.

Steele, J., J. Broadbent, and J. Kok. 2013. Perspectives on the contribution of lactic acid bacteria to cheese flavor development. Curr. Opin. Biotech. 24:135–141.

Tan, J., R. Li, Z.-T. Jiang, S.-H. Tang, Y. Wang, M. Shi, Y.-Q. Xiao, B. Jia, T.-X. Lu, and H. Wang. 2017. Synchronous front-face fluorescence spectroscopy for authentication of the adulteration of edible vegetable oil with refined used frying oil. Food Chem. 217:274–280.

Teófilo, R.F., J.P.A. Martins, and M.M.C. Ferreira. 2009. Sorting variables by using informative vectors as a strategy for feature selection in multivariate regression. J. Chemometr. 23:32–48.

Twomey, P.J. 2006. How to use difference plots in quantitative method comparison studies. Annal. Clinic. Biochem. 43:124–129.

van der Ven, C., H. Gruppen, D.B.A. de Bont, and A.G.J. Voragen. 2001. Reversed phase and size exclusion chromatography of milk protein hydrolysates: relation between elution from reversed phase column and apparent molecular weight distribution. Int. Dairy J. 11:83–92.

Walstra, P., T. J. Geurts, A. Noomen, A. Jellema, and M. A. J. S. Van Boekel. 2005. Chapter 1: Milk – composition, structure, and properties. Pages 1–24 in Dairy Technology: Principles of Milk Properties and Processes. Marcel Dekker Inc., New York, NY.

Wang, L.-L., Y.-W. Lin, X.-F. Wang, N. Xiao, Y.-D. Xu, H.-D. Li, and Q.-S. Xu. 2018. A selective review and comparison for interval variable selection in spectroscopic modeling. Chemom. Intell. Lab. Syst. 172:229–240.

Williams, P., and K. Norris. 2001. Near-infrared technology in the agricultural and food industries, 2nd ed. P. Williams and K. Norris, ed. AACC, St. Paul, MN

Wold, S., M. Sjöström, and L. Eriksson. 2001. PLS-regression: a basic tool of chemometrics. Chemom. Intell. Lab. Syst. 58:109–130.

Chapter 3 - Research objectives

This thesis focused on developing front-face fluorescence spectroscopy (FFFS)-based methods to quantify casein in fluid milk and detect protein leaks in permeate during ultrafiltration (UF). The specific objectives of the study are

- To develop and validate a quantification method for casein in raw milk using FFFS and principal component regression.
- To develop and validate a quantification method for casein content and casein to total protein ratio in fluid milk using FFFS and feature selections.
- To apply FFFS to identify the fluorophore of interest associated with protein leak, detect the presence of proteinaceous matters in UF permeates, and establish predictive models for protein leak in permeate streams.

Chapter 4 - Development and validation of a front-face fluorescence spectroscopy-based method to determine casein in raw milk[1]

Abstract

The objective of this study was to develop and validate a front-face fluorescence spectroscopy (FFFS)-based method for measuring casein in raw milk. Calibration samples (n_{cal} = 20) with different casein contents (0.36 - 3.7%) were prepared by mixing ultrafiltered retentate (2x concentrate) and permeate at different ratios. A principal component regression model was developed using the calibration fluorescence spectra, showing a useful prediction power based on a residual prediction deviation of 3.1. The calibration model was independently validated using 20 raw milk samples. The FFFS-based method showed a root mean square error of prediction of 0.15% and 6.7% relative prediction error. A larger sample size should be included in the future to further validate the method and potentially implement it for routine measurement of casein levels in raw milk.

Introduction

Spectroscopic methods combined with chemometrics have been used to provide rapid analysis of chemical, physical, and functional properties of dairy foods and dairy processes. Prior to cheese-making, manufacturers often standardize casein-to-fat ratio in cheese milk to optimize the production yield (Guinee et al., 2006). The standard casein measurements from Association of Analytical Chemists (AOAC) involve using hazardous reagents and require trained operators. Currently, milk casein quantification methods have been developed based on Fourier transform mid-infrared spectroscopy (FT-MIR) (McDermott et al., 2016) and near-infrared spectroscopy

[1] Published in International Dairy Journal (93): 81-84

(NIR) (Barbano and Dellavalle, 1987). The NIR method was based on the indirect measurement of casein by taking the difference between total protein and serum phase protein. The study of McDermott et al. (2016) applied partial least square regression (PLSR) to FT-MIR spectra collected for raw milk to predict casein content, as measured by a high-performance liquid chromatography (HPLC)-based method.

Milk contains several intrinsic fluorophores that are suitable for fluorometric analysis. Front-face fluorescence spectroscopy (FFFS) has shown the potential to directly measure turbid samples for classification and quantification of dairy food composition. Tryptophan, as a compound of interest, is commonly used for method development, as its emission spectra can be well-correlated with protein content, acid coagulation, and heat treatment in milk (Andersen and Mortensen, 2008; Birlouez-Aragon et al., 1998). However, FFFS has not been utilized as a tool to measure casein in raw milk but may provide an alternative casein quantification method for the cheese industry. The objective of this study was to develop and validate a quantification method for casein in raw milk using FFFS and principal component regression.

Materials and methods

Sample preparation for calibration set

Two batches of commingled raw milk were procured from the Kansas State University Dairy Cattle Teaching and Research Unit (Manhattan, KS, USA) on two different random days. The raw skim milk was produced using a bench-top cream separator (Motor Sich-100, Ukraine). For each batch, ultrafiltration of raw skim milk was carried out at 25ºC using a bench-top plate and frame system (SmartFlow Technologies, Apex, NC, USA) equipped with a 10-KDa cut-off polyethersulfone membrane (Hannifin Corp., Oxnard, CA, USA). Approximately, 2x retentate was produced at a constant transmembrane pressure of 207 KPa. Subsequently, the retentate and permeate were combined to make 10 calibration samples with different casein levels. In total, 20 calibration samples were prepared with a uniform distribution of casein content ranging from 0.36 to 3.7% (w/w) and whey protein-to-casein ratio ranging from 0.22 to 0.27.

Reference measurement of casein

Total nitrogen and non-protein nitrogen of retentate and permeate were analyzed using AOAC (2016) standard methods (990.20 and 990.21, respectively). Due to the high protein content in the retentate, the non-casein nitrogen was measured using the method described in Zhang and Metzger (2011). Casein content was obtained from the difference between total nitrogen and non-casein nitrogen multiplied by 6.38.

FFFS measurement of casein

Sample preparation prior to acquisition of fluorescence spectra involved a patent-pending precipitation of caseins. Preliminary experiments have revealed spectral difference between casein-precipitated (pH = 4.6) raw milk and raw milk at natural pH (data not shown). Prior to FFFS, 7 mL of calibration sample was taken in a 10-mL test tube and mixed with 0.6 mL of 10 % acetic acid (Certified ACS, Fisher Scientific, Hampton, NH, USA) to ensure a pH of 4.6 ± 0.05. The mixture was vortexed for 15 s and transferred immediately into a Quartz cuvette (Starna Cells Inc., Atascadero, CA, USA), ensuring no phase separation. Tryptophan emission spectra were immediately acquired using a spectrofluorimeter fitted with a 1% attenuator (LS-55; Perkin Elmer, Waltham, MA, USA) at an excitation wavelength 280 nm and an emission scan from 300 to 440 nm. Triplicate measurements on freshly precipitated milk samples were performed at 25ºC and averaged to improve signal-to-noise ratio.

Construction of a calibration model

Pre-processing tools such as normalization and derivation are commonly used to reduce drift noise and reveal spectral overlays prior to the model development. In this study, the first derivative of the tryptophan emission spectra was obtained using the Savitzky-Golay algorithm with 9-point neighbor values (Savitzky and Golay, 1964). No light-scattering effect was observed from preliminary experiments, so the scattering subtraction was not conducted as part of preprocessing. Supervised prediction models were developed from the first derivative spectra using principal component regression (PCR) and evaluated with leave-one-out cross-validation. The model performance was evaluated in terms of root mean square error of cross-validation (RMSECV; Equation 1). Residual prediction deviation (RPD) was calculated from RMSECV divided by the sample standard deviation and used as a parameter to estimate the model prediction power. Statistical model building and evaluation were conducted using RStudio (version 1.1.442; RStudio, Boston, MA) with the Caret package (Kuhn, 2008).

Independent validation of the model

To validate the PCR-based calibration model, raw milk samples were randomly collected from 20 individual cows ($n_{val} = 20$) from the same dairy farm, and casein content was measured by the reference method as described above. Subsequently, tryptophan FFFS spectra were also collected from the validation samples as described above and preprocessed following the same procedure used for calibration samples. The accuracy of the FFFS-based method was evaluated in terms of the mean bias and root mean square error of prediction (RMSEP; Equation 1).

$$RMSECV\ or\ RMSEP = \sqrt{\frac{\sum_{j=1}^{n}(y_{j-}\ \hat{y}_j)^2}{n}} \qquad (1)$$

Where y is the predicted value from cross-validation or independent validation of the j^{th} sample, $\hat{y}$ is the reference value of the j^{th} sample, and n is sample size.

Relative prediction error (RPE) was also calculated dividing the RMSEP by the average casein content obtained from the reference method (Amamcharla and Panigrahi, 2010). The range error ratio (RER) was calculated by dividing the range of reference casein content by the RMSEP. Graphical representation was prepared by plotting the difference and the averaged casein content measured by the reference and FFFS-based methods. This difference plot approach allowed the comparison of bias and error between the two methods and determine outliers from the measurements (Twomey, 2006).

Results and discussion

Tryptophan fluorescence spectra of casein in calibration samples

Figure 4-1A shows representative tryptophan fluorescence spectra of low (0.36%) and high (3.7%) casein samples from the calibration set. The remaining calibration samples also produced similar tryptophan spectra ranged between the low and high-casein samples (data not shown). The tryptophan emission maximum was observed at 338.5 ± 1 nm for all the samples. Shaikh & O'Donnell (2017) reviewed tryptophan fluorescence in milk and reported emission maxima around 340 nm with excitation of 290 nm. Figure 4-1B shows the first derivative spectra of low and high-casein samples from the calibration set. More spectral variations were observed around emission wavelengths of 313 and 363 nm. The derivation also smoothed the spectra and unified the fluorescence intensity scale and was subsequently used for calibration model development.

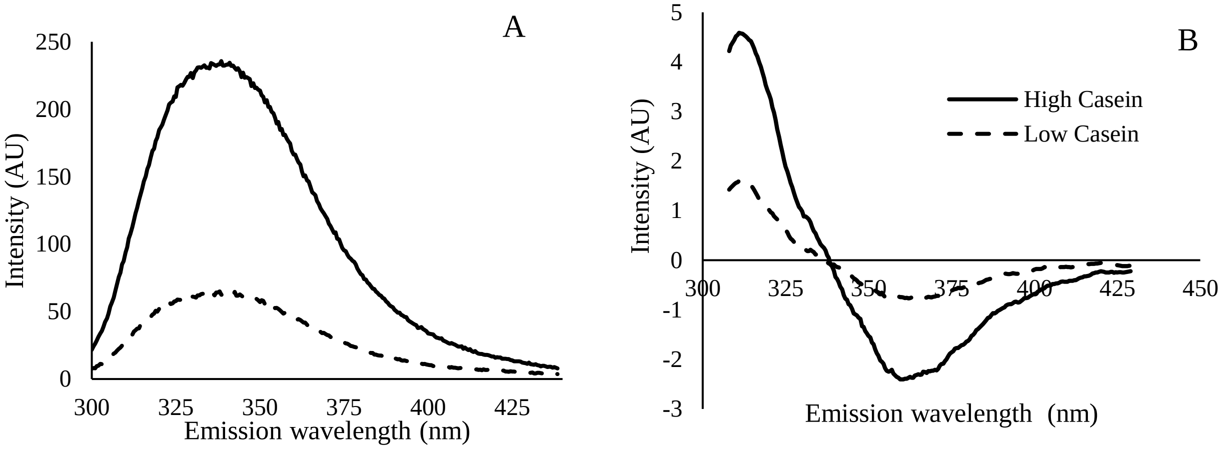


Figure 4-1 A: Representative front-face tryptophan fluorescence spectra (excitation = 280 nm) of low (dashed line) and high-casein (solid line) calibration samples; B: first derivative of spectra in Fig. 1A. AU = arbitrary unit.

Calibration model development

Figure 4-2 shows the scatter plot between the casein content obtained from the FFFS-based method and the reference method. For the reference method, the duplicated casein measurements showed an average coefficient of variations of 3.5%, providing reliable model reference for the calibration. The optimum number of principal components (n_{pc} = 2) was chosen based on the lowest RMSECV (= 0.35) of the calibration model. The coefficient of determination (R^2) was found to be 0.89 and the slope for the predicted vs reference casein values was 0.91, showing an increasing negative bias towards the high-casein samples. The calibration samples were evenly distributed with equal sample size for different casein content. This approach was used to minimize the “Dunne effect” which leads the overestimation of the lower range and underestimation of the higher range in calibrations (Williams and Norris, 2001). The intercept of the scatter plot was 0.18, indicating that the it may result in more measuring error for low-casein samples. The difference in casein precipitations between the high and low casein samples may lead to spectral variation during measurement and account for the measuring errors. The prediction power of the model was evaluated by RPD, and, according to Williams and Norris (2001), an RPD of 3.1 for the model indicates a useful prediction.

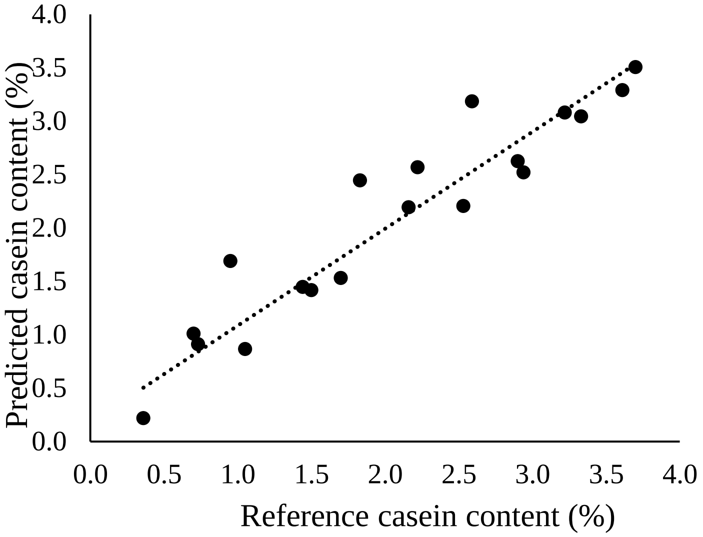


Figure 4-2 Principal component regression calibration model: reference versus predicted casein content (%). Number of principal components was 2, coefficient of determination was 0.89, slope is 0.91, intercept is 0.18, root mean square error of crossvalidation is 0.35, residual prediction deviation is 3.1.

Independent validation

According to the difference plot shown in Figure 4-3, 19 out of 20 validation samples fall within the 95% confidence interval, with only 1 sample falling beyond the lower confidence interval. This sample had a reference casein value of 2.34% (w/w) and predicted value of 2.01% (w/w). Due to the small cow-to-cow variation in the casein content among validation samples, R^2 may not fully evaluate the model performance (Altman and Bland, 1983). In this study, the mean casein content bias between the reference and FFFS-based methods was 0.01%, indicating a slight underestimation of the FFFS-based method. The RMSEP was 0.15% and the corresponding RPE of the FFFS-based method was 6.7%.

McDermott et al. (2016) conducted FT-MIR measurement of casein in milk, and the external validation of the casein prediction resulted in an RPE of 13% using HPLC-based reference values. Hewavitharana and Brakel (1997) developed a FT-IR method for casein in raw milk using PLSR and PCR and results showed an RPE between 2.3 to 4.3%. In general, an RPE less than 5% is desirable for accurate analysis, so the FFFS-based method can still be improved to fulfill the accuracy requirement. For the present study, the RER of the validation was 8.4. According to Williams and Norris (2001), RER describes the practicality of such models, and a value above 4.0 is acceptable for sample screening purposes. Therefore, the FFFS-based method developed in this study still has practical use to serve the dairy industry.

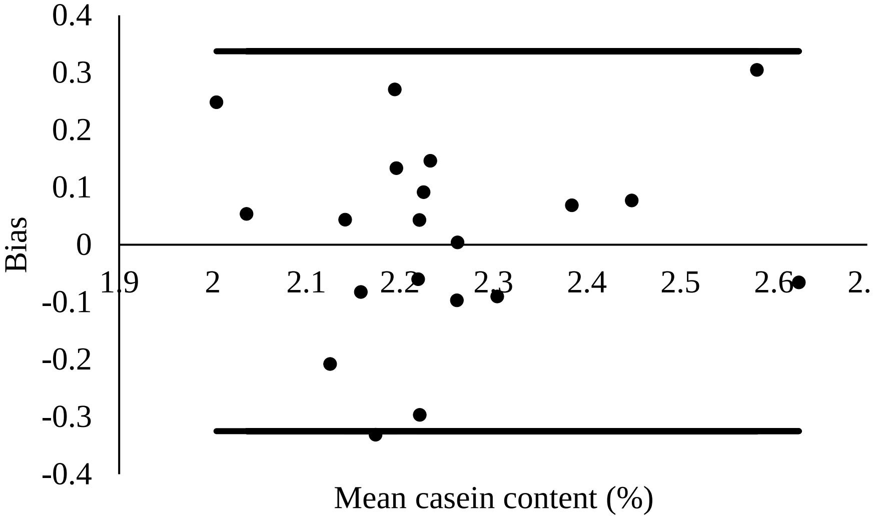


Figure 4-3 Difference plot showing the bias between reference and predicted casein contents (n_{val} = 20) against the mean casein content (%). The black bars indicate the 95% confidence intervals. Mean bias: 0.01%, root mean square error of prediction: 0.15%, relative prediction error: 6.7%, range error ratio: 8.4.

Conclusions

This study validated a FFFS-based method to quantify casein in raw milk. Tryptophan front-face fluorescence spectra of casein suspension obtained from acidifying raw milk to pH 4.6 was used to develop a calibration model. The model showed practical prediction power and an independent validation confirmed the industry feasibility. For future studies, the FFFS-based method should be validated with a larger sample size including more sources of variation such as season, feeding, and cow breed. The method will be further adapted to a portable fluorescence device and potentially implemented for routine measurement of casein in raw milk.

References

Amamcharla, J. K., S. Panigrah 2010. Simultaneous prediction of acetic acid/ethanol concentrations in their binary mixtures using metalloporphyrin based opto-electronic nose for meat safety applications. Sens. Instrum. Food Qual. and Saf. 4: 51-60.

AOAC. 2016. Official methods of analysis (20th ed.). Association of Official Analytical Chemists. Gaithersburg, MD, USA: AOAC International.

Altman, D. G. and J.M. Bland. 1983. Measurement in medicine: the analysis of method comparison studies. J. Royal Stat. Soc. 32: 307–317.

Andersen, C. M. and G. Mortensen. 2008. Fluorescence spectroscopy: a rapid tool for analyzing dairy products. J. Ag. Food Chem. 56: 720–729.

Barbano, D. M. and M.E. Dellavalle. 1987. Rapid method for determination of milk casein content by infrared analysis. J. Dairy Sci. 70: 1524–1528.

Birlouez-Aragon, I., M. Nicolas, A. Metais, N. Marchond, J. Grenier, and D. Calvo. 1998. A rapid fluorimetric method to estimate the heat treatment of liquid milk. Int. Dairy J. 8: 771-777.

Guinee, T. P., B.T. O'Kennedy, and P.M. Kelly. 2006. Effect of milk protein standardization using different methods on the composition and yields of cheddar cheese. J. Dairy Sci. 89: 468–482.

Hewavitharana, A. K. and B. van Brakel, 1997. Fourier transform infrared spectrometric method for the rapid determination of casein in raw milk. Analyst, 122: 701-704.

Kuhn, M. 2008. Caret package. J. Stats. Softw. 28: 1-26.

McDermott, A., G. Visentin, M. De Marchi, D.P. Berry, M.A. Fenelon, P.M. O'Connor, … and S. McParland. 2016. Prediction of individual milk proteins including free amino acids in bovine milk using mid-infrared spectroscopy and their correlations with milk processing characteristics. J. Dairy Sci. 99: 3171–3182

Savitzky, A. and M.J.E Golay. 1964. Smoothing and differentiation of data by simplified least squares procedures. Anal. Chem. 36: 1627–1639.

Shaikh, S., and C. O'Donnell. 2017. Applications of fluorescence spectroscopy in dairy processing: a review. Curr. Opin. Food Sci. 17: 16-24.

Twomey, P. J. 2006. How to use difference plots in quantitative method comparison studies. Ann. Clin. Biochem. 43: 124–129.

Williams, P., and K. Norris. 2001. Near-infrared technology in the agricultural and food industries, 2nd ed. P. Williams and K. Norris, ed. AACC, St. Paul, MN

Zhang, H., and L.E. Metzger. 2011. Noncasein nitrogen analysis of ultrafiltration and microfiltration retentate. J. Dairy Sci. 94: 2118–2125.

Chapter 5 - Quantification of casein and casein-to-crude-protein ratio in fluid milk by front-face fluorescence spectroscopy combined with chemometrics[2]

Abstract

Casein in fluid milk determines cheese yield and impacts cheese quality. Traditional methods of measuring casein in milk involves lengthy sample preparations with labor-intensive nitrogen-based protein quantifications. Previous studies have applied front-face fluorescence spectroscopy (FFFS) combined with chemometrics to measure physical, chemical, and function properties of dairy foods. The objective of this study was to quantify casein and casein-to-crude-protein ratio (CN/CP) in fluid milk using FFFS and chemometrics. Calibration samples were constructed by mixing of microfiltration and ultrafiltration retentate and permeate in different ratio to obtain different concentrations of casein and CN/CP ratio. Partial least squares regression (PLSR) and elastic net regression models were developed for casein and CN/CP prediction in fluid milk using FFFS tryptophan emission spectra and reference casein contents. A set of 20 validation samples including raw, skim, and UF milk was used to optimize and validate the model performance. Another independent set of 20 test samples including raw, skim, and UF milk samples were externally tested in terms of root mean square error of prediction (RMSEP), residual prediction deviation (RPD), and relative prediction error (RPE). The RMSEP for casein content quantification in raw, skim, and UF milk ranged from 0.12 to 0.13% with RPD ranged

[2] Submitted for publication: Food and Bioprocess Technology

from 3.2 to 3.4, indicating a practical model performance for quality control purposes. For CN/CP quantification, PLSR developed useful prediction model with RMSEP of 0.024, RPD of 1.5, and RPE of 3.0. This FFFS-based method can be implemented to a routine quality control tool in the dairy industry, providing rapid quantifications of casein content and CN/CP in fluid milk intended for cheese manufacturing.

Introduction

Improving cheese yield is a constant pursuit among cheese manufacturers. Dairy food scientists have identified multiple factors that can influence cheese yield. Milk compositions, namely the amount of casein and fat, have been highlighted in multiple studies as indicators that determine cheese yield (Barbano and Sherbon, 1984; Emmons and Modler, 2010). Several cheese yield determinants included curd firmness, syneresis rate, and moisture retention. These qualities of the cheese curd have been partially linked back to the casein content in the cheese milk (Cipolat-Gotet, 2013). Standardization of cheese milk therefore became a common practice in cheese manufacturing. By adjusting the casein-to-fat ratio, the standardized milk can maximize cheese yield without losing excessive fat and casein into whey (Lucey and Kelly, 1994). Moreover, with the advancement in membrane processing technologies, cheese can be produced from ultrafiltered (UF) and microfiltered (MF) milk. UF and MF milk contains more casein than regular cheese milk which can increase the cheese yield and improve vat utilizations (Kumar et al., 2013).

The current standard method of casein measurement involves isoelectrically precipitating casein at pH of 4.6 in milk and separating casein from the non-casein fraction by filtration. According to the AOAC standard method (990.20 and 990.21). Casein content can either be directly measured on the isolated casein solids or indirectly calculated as the difference between total protein and non-casein proteins by using a Kjeldahl-based method. This Kjeldahl-based method resulted in good repeatability and reproducibility, serving as the industry standard method for casein quantification since 1938 (Rowland, 1938; Lynch et al., 1998). However, this quantification process is laborious, time-consuming, and uses multiple hazardous chemical reagents. Dairy food researchers have proposed two general alternative approaches to measure casein based on separation techniques and infrared spectroscopies. High performance liquid chromatography separations based on reverse phase, gel permeation, and size exclusion have been developed to fractionate β-casein, κ-casein, αs_1-casein, and αs_2-casein (Dimenna and Segall, 1981; van der Ven et al., 2001; Bonfatti et al., 2008). Similarly, capillary electrophoresis-based method was developed to quantify whey protein and casein in milk (Recio and Olieman, 1996). While the primary goal of these methods was for protein separation, quantification of casein can also be achieved by using appropriate standards. Infrared spectroscopic methods of casein quantification have been developed with the help of multivariate statistical models. One early attempt of near infrared (NIR) measurement of casein was based on the indirect approach by taking the difference between total protein and serum phase protein measured by NIR-based protein measurement (Barbano and Dellavalle, 1987). With the advancement of Fourier-transform infrared spectroscopy (FTIR), Hewavitharana and van Brakel (1997) and Luginbuhl (2002) both developed and validated casein quantification with FTIR and multivariate statistical models. Foss Electric has also implemented calibration options for measuring casein in fluid milk using its FTIR model “MilkoScan FT 120” (Foss Electric, 1997). These spectroscopic methods aimed to reduce labor investment and analysis time for casein quantification in fluid milk.

There are multiple intrinsic fluorophores in milk which are suitable for fluorescence spectroscopic analysis. Front-face fluorescence spectroscopy (FFFS) is known for its sensitivity and ability to analyze turbid samples. Tryptophan, a commonly studied fluorophore in milk, can measure milk coagulation, degrees of heat treatment, and dairy powder solubility during storage (Herbert et al., 1999; Kulmyrzaev et al, 2005; Babu and Amamcharla, 2018). Herbert et al. (1999) studied tryptophan emission spectra and characterize acid coagulation process in milk. The study indicated that in acidic conditions, casein in milk yielded fluorescence spectral difference compared to milk with native pH. Ma et al. (2019) measured precipitated casein tryptophan fluorescence in raw milk and established multivariate calibration models to quantify casein. This study extends the previous study aiming to measure casein content and casein-to-crude-protein ratio (CN/CP) in raw, skim, and ultrafiltrated (UF) milk using FFFS and chemometrics.

Materials and methods

Experimental approach

Calibration samples were constructed by mixing permeates and retentates obtained from UF and MF of pasteurized skim milk in different ratios to obtain different concentrations of casein and CN/CP ratio. Multivariate calibration models were developed using tryptophan emission spectra and reference values for casein content and CN/CP based on the Kjeldahl method. Validation and test samples from raw, skim, and UF milk optimized and tested the final model performance for casein content and CN/CP measurements. Detailed methods are described in the following sections.

Calibration, validation, and test sample

UF retentate and permeate (about 5X concentrated) made from one lot of pasteurized skim milk were donated by a commercial milk protein concentrate manufacturer in the United States. MF retentate and permeate (about 3X concentrated) from one lot of pasteurized skim milk were donated by the Southeast Dairy Foods Research Center (Raleigh, NC). Both UF and MF milk fractions arrived under refrigerated conditions and were analyzed for casein content based on the Kjeldahl method (AOAC 990.20 and 990.21).

After measuring the casein content in UF and MF permeate and retentate, the UF and MF retentates were diluted with various amount of UF and MF permeates to vary casein content and CN/CP in calibration samples. Consequently, 30 calibration samples (n_{cal} = 30) were prepared with casein contents ranged from 1.21 to 4.45% and CN/CP ranged from 0.66 to 0.88 and used for calibration model development (Table 5-1).

For validation and test samples, 10 pasteurized skim milk and 10 UF milk with different production days were purchased from local supermarkets. Additional 20 raw milk samples from individual cows were randomly collected from Kansas State University Dairy Cattle Teaching and Research Unit (Manhattan, KS). The validation and test samples ($n_{val/test}$ = 40) were stored at 5 °C until further analysis.

Table 5-1 Mean (range) protein fractions of calibration samples (n = 30), ultrafiltered milk (n = 10), pasteurized skim milk (n = 10), and raw milk (n = 20)

	Calibration samples	Validation and test samples		
		Raw milk	Pasteurized Skim milk	Ultrafiltered milk
Crude protein (%)	3.29 (1.82 – 5.17)	2.91 (2.57 – 3.34)	3.16 (2.75 – 3.42)	3.77 (3.54 – 3.93)
Casein (%)	2.57 (1.21 – 4.45)	2.25 (1.88 – 2.66)	2.37 (2.03 – 2.64)	3.17 (2.90 – 3.34)
Non-protein nitrogen (% protein-equivalent)	0.14 (0.10 – 0.26)	0.12 (0.09 – 0.14)	0.12 (0.10 – 0.14)	0.17 (0.14 – 0.20)

Casein-to-crude-protein ratio	0.77 (0.66 – 0.88)	0.77 (0.73 – 0.80)	0.76 (0.70 – 0.79)	0.84 (0.82 – 0.86)

Reference measurement of casein and calibration sample preparation

Crude protein (CP) and non-protein nitrogen (NPN) of UF and MF retentate, permeate, and validation and test milk samples were analyzed by AOAC (2016) standard methods (990.20 and 990.21, respectively). Due to the high protein content in the MF and UF retentate, the non-casein nitrogen (NCN) was measured according to Zhang and Metzger (2011). Casein content was obtained from the difference between CP and NCN and multiplied by 6.38. CN/CP ratio was calculated using the casein content divided by the crude protein to represent the proportion of casein in relationship to total protein of the milk sample.

Tryptophan fluorescence collection for calibration and validation samples

Based on preliminary studies, completely precipitating casein at pH of 4.6 yielded distinctive spectra compared to the rest of the pH-adjusted and native samples. The FFFS spectral collection was achieved according to Ma et al. (2019). Prior to FFFS measurement, 7 mL of sample was taken in a 10-mL test tube and mixed with 0.6 mL of 10 % acetic acid (Fisher Scientific, Hampton, NH) to ensure a pH of 4.60 ± 0.05. The mixture was vortexed for 15 s and transferred immediately into a Quartz cuvette (Starna Cells Inc., Atascadero, CA, USA) ensuring no phase separation. Tryptophan emission spectra were immediately acquired using a spectrofluorimeter fitted with a 1% attenuator (LS-55; Perkin Elmer, Waltham, MA, USA) at an excitation wavelength 280 nm and an emission scan from 300 to 440 nm with a scan speed of 300 nm/min. Triplicate measurements on freshly precipitated calibration samples were performed at 25ºC and averaged improve signal-to-noise ratio. In total, tryptophan fluorescence spectra were collected on 30 calibration and 40 validation and test samples for the chemometric model development.

Chemometric model development, optimization, and validation

Developing chemometric models involves optimization and test of the finalized models (Bevilacqua et al., 2017). A detailed model development approach followed in this study can be found in Figure 5-1. Calibration models were developed using FFFS tryptophan spectra and reference casein content and CN/CP of calibration samples (n_{cal} = 30). The validation and test samples ($n_{val/test}$ = 40) were randomly partitioned into a validation set (n_{val} = 20) and a test set (n_{test} = 20). A summary of the casein content and CN/CP of the validation and test sets can be found in Table 5-1. The validation set was used to validate and optimize the preliminary models, while the test set was used to evaluate the optimized model performance for casein content and CN/CP quantifications. The quantification results from the test set provide estimation of the future performance of the developed model on new data.

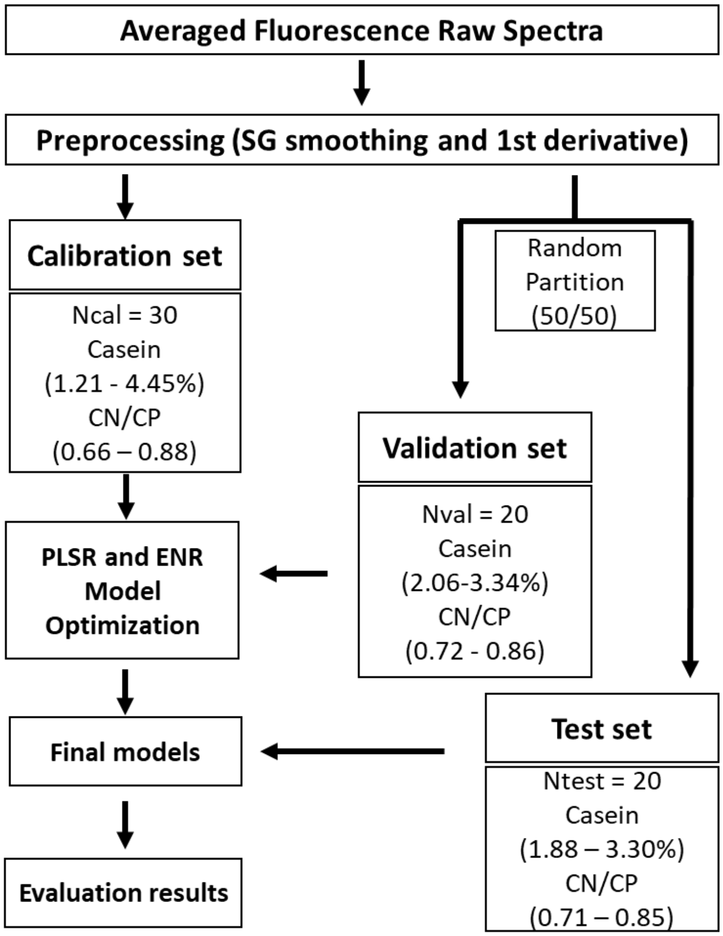


Figure 5-1 Chemometric model development overview with casein (range) and casein-to-crude-protein ratio (CN/CP; range). SG = Savitzky-Golay; Ncal = number of calibration samples, Nval = number of validation samples, Ntest = number of test samples; PLSR = Partial least square regression; ENR = Elastic net regression; RMSE = Root mean square error, RPD = Residual prediction deviation; RPE = Relative prediction error.

Spectral preprocessing and construction of calibration models

Preprocessing tools such as normalization, derivation, and smoothing are commonly used to reduce drift noise and reveal spectral overlays prior to the model development (Brown et al. 2000). In this study, raw fluorescence spectra were transformed using the Savitzky-Golay smoothing (SG-S) and first derivative (SG-1st) algorithms with 9-point neighbor values to reduce the spectral noise from directly measuring turbid milk samples and reveal additional spectral information (Savitzky and Golay, 1964).

Supervised prediction models were developed using partial least square regression (PLSR) and elastic net regression (ENR). In chemometrics, PLSR is a popular method for relating two data matrices using a linear multivariate model capable of handling large number variables with noise and collinearity (Wold et al., 2001). ENR is a type of panelized linear regression with the capability to eliminate and shrink variable contributions in multivariate models (Chen et al., 2014). Chemometric researchers have applied both PLSR and ENR to model spectral data for their ability to handle large number of predictors (Filzmoser et al. 2012). The model input consisted either the smoothed or first derivative of the tryptophan emission spectra, and casein content and CN/CP were predicted independently using the preprocessed spectra. In this study, up to 15 latent variables were considered in the initial model development of PLSR and considered as the model optimization parameter. For ENR, elastic net parameter (alpha) and regularization parameter (lambda) were considered as optimization parameters. In this study, an increment of 0.1 was used for alpha optimization using leave-one-out crossvalidation to find the best performing lambda. A total of 10 models from ENR was produced from the calibration step and they were later optimized by the validation set for the optimal alpha value.

Model optimizations

The established calibration models were optimized by predicting the validation set. Model performance was evaluated in terms of root mean square error of validation (RMSEV; Equation 1) and coefficient of determination (R^2) between the reference and predicted values. Optimization parameters for PLSR (number of latent variables) and ENR (alpha, and lambda values) were selected based on the lowest RMSEV. The R^2 evaluated the linearity of the model prediction to the reference values, and calibration transfer based on linear models was used to correct the estimated bias from the preliminary PLSR or ENR predictions. The optimal parameters of PLSR and ENR were recorded and used for the final test set predictions.

Model performance evaluation

The model evaluation was achieved by predicting the test set with the finalized PLSR and ENR models. The final model performance was evaluated by the root mean square error of prediction (RMSEP), showing the difference between predicted values and the reference values. Residual prediction deviation (RPD) was calculated from RMSEP divided by the standard deviation of reference values and used as a parameter to estimate model's prediction power. Relative prediction error (RPE) were calculated dividing the average reference values by the RMSEP to evaluate the relative error of the prediction to the reference method. Spectral preprocessing, statistical model building and evaluation were conducted using an in-house program developed in RStudio (version 1.1.442; RStudio, Boston, MA) with the Caret, pls, and glmnet packages (Kuhn, 2008; Friedman et al., 2010; Mevik and Wehrens, 2007).

Results and discussion

Reference Casein and CN/CP Measurements

In general, average casein content in milk is reported between 2.6 to 2.8 % (Walstra and Jenness, 1984; Fox and McSweeney, 1998). However, the reported average casein content in milk can be impacted by season, diet, and genetic variations of dairy cows (Lin et al., 2017). Lin et al. reported the casein content in comingled pasteurized skim milk from Holstein Friesian over one year and found that casein content could range from 2.61 to 3.02%. Chen et al. (2014) monitored comingled raw milk casein content ranged from 2.08 to 2.52 % for one year. With the recent popularity of high-protein beverages, UF milk as a consumer product has entered the market. According to a high-protein milk application developed by Ur-Rehman et al. (2017), casein in UF and delactosed milk can range from 2 to 8% during production, and for the finished product, CP content can range from 4.9% to 5.2%. Even though casein content was not specified for the finished product, the casein content was assumed to be less than the reported total protein content. Table 5-1 summarized the protein fractions of UF milk obtained from this study. For the 10 commercial UF milk samples obtained, the casein content ranged from 2.90 to 3.34% with CP ranged from 3.54 to 3.93%. For the commercial pasteurized skim milk samples, the casein content ranged from 2.03 to 2.64% with CP ranged from 2.75 to 3.42%. Casein content for raw milk samples in this study had a range from 1.88 to 2.66% with CP ranged from 2.57 to 3.39%. The casein content variation observed from this study suggested that the calibration range for measuring casein need to cover the casein range for raw, skim, and UF milk. In Table 5-1, the casein content of the calibration samples ranged from 1.21 to 4.45%, providing a sufficient calibration range for measuring casein in raw, skim, and UF milk.

The casein content variation in milk can also cause CN/CP variations. Lin et al (2017). reported the CN/CP for pasteurized skim milk from Holstein Friesian cows ranged from 0.75 to 0.81 during the one-year observation. According to Schaar et al. (1985), genetic variants affected κ-casein and β-lactoglobulin synthesis during lactation which led to variations in casein number (CN/CP x 100) and cheese compositions (Lundén et al., 1997). The CN/CP variation was observed in this study. The CN/CP variation in UF milk ranged from 0.82 to 0.86, pasteurized skim milk ranged from 0.71 to 0.79, and raw milk ranged from 0.73 to 0.80 (Table 1). The variation in milk samples requires a set of calibration samples that covers the target CN/CP range. Table 5-1 showed that the calibration sample range was obtained by mixing various amount of UF and MF retentate and permeate, which produced CN/CP ranged from 0.66 to 0.88.

Tryptophan Fluorescence Emission Spectra of Calibration Samples

Tryptophan fluorescence emission spectra of acid-precipitated calibration samples were collected using FFFS. The emission maxima (λmax) of the calibration samples (ncal = 30) ranged from 338 to 341 nm. According to Andersen and Mortensen (2008), emission maxima around 340 nm confirms the fluorophore to be tryptophan. The calibration set varied in casein contents and CN/CP (Table 1) and led to tryptophan emission intensity differences. Figure 5-2A shows the casein content in sample I (1.6%) is less than sample II (2.1%). The NCN level between samples I and II was also different showing 0.9% and 0.4%, respectively. The CP level for samples I and II end up being very similar at 2.5%. In Figure 5-2B, the tryptophan emission spectra of samples I and II appeared to have large fluorescence emission intensity difference with an emission maximal (λmax) increase of 23.3%. On the other hand, a similar amount of casein present in samples II and III (2.1%), but their CP contents were different because the NCN of sample III is higher than sample II (1.3% and 0.4%, respectively). The emission spectra of samples II and III appeared to be similar with 2.3% change in λmax. Samples III and IV had similar amount of CP (3.4%) but different casein content (2.1 and 2.7%, respectively). The difference in casein content is again highlighted in the change of emission spectra shown in Figure 5-2B. The emission spectral difference among samples I, II, III, and IV illustrated that the tryptophan emission spectra of the acid-precipitated casein are more sensitive to casein content as oppose to CP content.

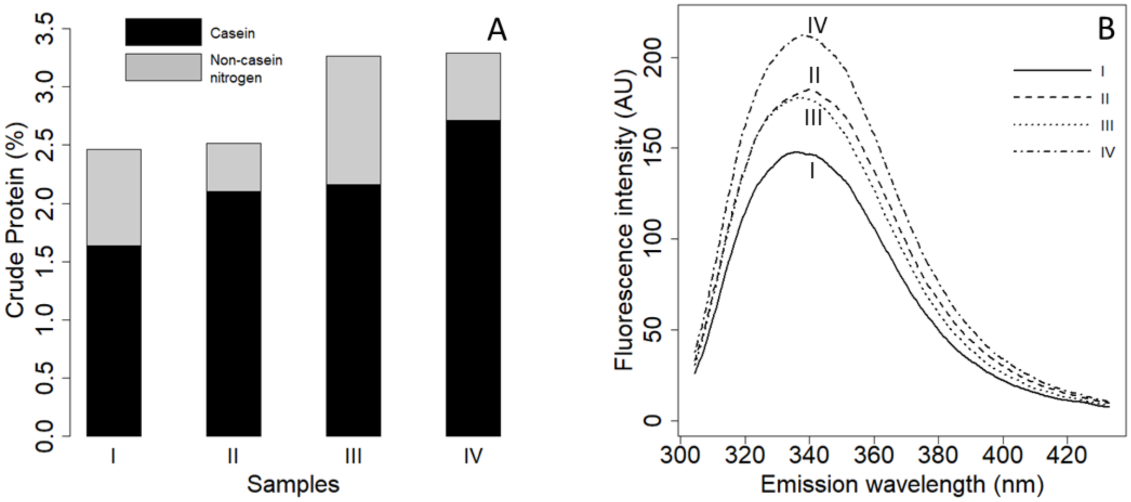


Figure 5-2 A: protein fractions (casein and non-casein nitrogen) of 3 representative calibration samples (I, II, III, and IV). B: tryptophan emission spectra of the corresponding samples.

At pH of 4.6, casein reaches its isoelectric point and aggregates in the milk dispersion system while the serum phase remains as a transparent liquid. The tryptophan-containing casein aggregates can absorb excitation light (280nm) and emit fluorescence at 300 to 400 nm (Herbert et al., 1999). In the meantime, the casein aggregates have surface protuberances, which could randomly scatter the excitation light (McMahon et al., 2009). The scattered excitation light may be again absorbed and emit more fluorescence by the casein aggregates. In the serum phase, tryptophan-containing whey proteins, peptides, and free amino acids will also absorb the excitation light and emit fluorescence (Birlouez-Aragon et al., 2002). However, due to the low optical density of the serum phase, less scattering effects could occurr. With the scattering of casein aggregate, tryptophan fluorescence observed in Figure 5-2 may reflect the change in casein content more than the change of crude protein content in a given milk system at pH of 4.6.

Calibration model development and optimization

PLSR and ENR models were constructed using the acid-precipitated casein tryptophan emission spectra as inputs. Since the calibration samples were lab-constructed, a validation set was necessary to ensure the model's validity on real milk samples. For PLSR model validation and optimization, the number of latent variables determined the model performance in terms of RMSEV and R^2 (Wold, 2001).

Model optimization for PLSR

A typical latent variable selection process is shown in Figure 5-3A for PLSR prediction of casein content using the SG-S preprocessing technique. The RMSEV (0.66%) showed minimal value with 2 latent variables, meaning the lowest prediction error for casein content. The R^2 (0.90) also had the highest value with 2 latent variables in the model, so the PLSR model with 2 latent variables was selected for the final prediction model. However, it is observed that the RMSEV from the validation was high, resulting in an RPE of 26.3%. The prediction of the validation set was visualized in Figure 5-3B, and both proportional and constant errors were observed with a slope of 0.46 and intercept of 1.14. Considering the high R^2 obtained from the validation, a linear model was applied to correct the estimation bias in the validation set. This approach of linear model correction is known as calibration transfer. These calibration transfer methods are more commonly seen in near infrared (NIR) spectroscopy model development to account for instrument signal drifts (Bouveresse et al., 1999; Liu et al., 2016). Less common is the calibration transfer method used on the same instrument. However, in this study, the calibration samples were lab-constructed samples. Although generated from milk, the calibration samples may yield different fluorescence intensity due to possible variations in total solids and mineral contents. The validation set was able to capture the high estimation bias and correct it as part of the model development process. The rest of PLSR models was optimized using the same approach and corrected with estimation bias using linear regression.

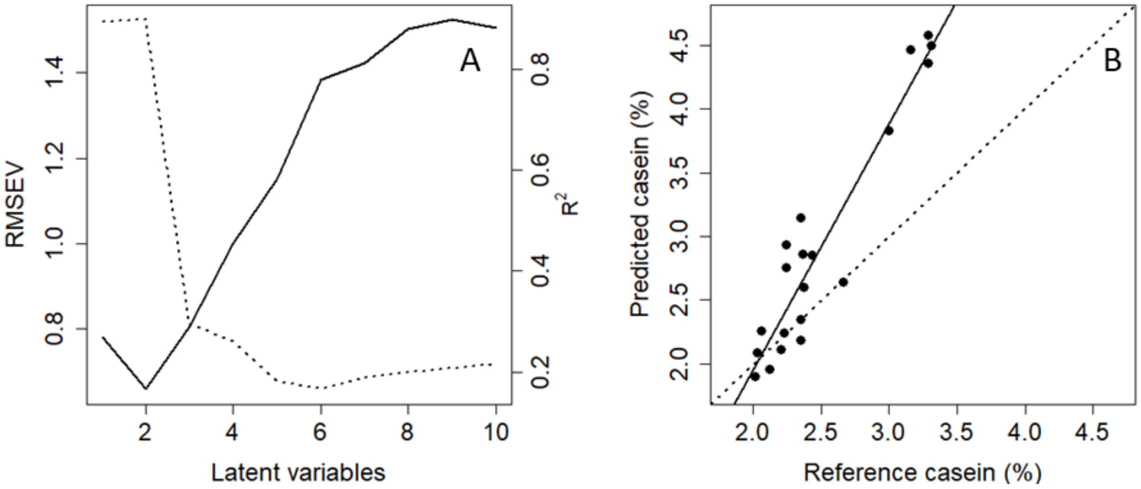


Figure 5-3 A: example of parameter optimization for partial least square regression (PLSR) for latent variable selections using the validation set. The solid line represents the change in root mean square error of validation (RMSEV) and the dotted line represents the represents the change in coefficient of determination (R^2). B: reference vs predicted casein (%) of the validation set using PLSR. The solid line represents the least-square fit of the scatter plot and the dot line presents ideal prediction target (x = y).

Model optimization for ENR

To optimize the ENR model optimization, the optimal alpha was selected using the lowest RMSEV generated from the validation set. In Figure 5-4A, a representative elastic net parameter selection process is shown for the casein content prediction using SG-S as the preprocessing step. The lowest RMSEV observed was 0.64% when alpha equals to 0.9. When alpha equals 0.1, R^2 reached the highest value of 0.9. When different alpha values were observed for the optimal RMSEV and R^2, the RMSEV was chosen as the evaluating criterium because it judges the true prediction power of the model (Geladi, 2002). Therefore, an alpha value of 0.9 was chosen to optimize the ENR model. Similar to the PLSR models, linear model correction was used to adjust the biased estimation plotted in Figure 5-4B for ENR predictions.

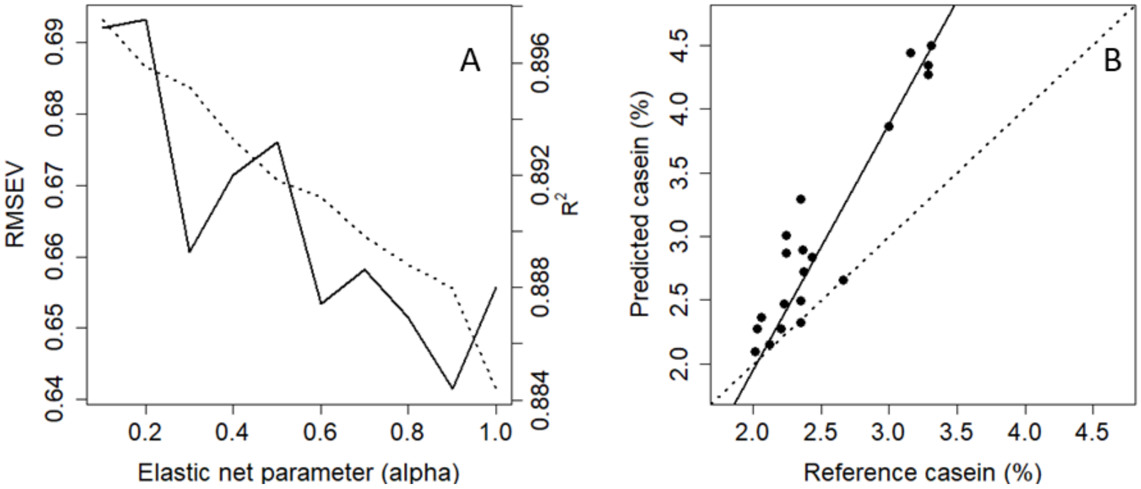


Figure 5-4 A: example of parameter optimization for elastic net regression (ENR)for elastic net parameter selections using the validation set. The solid line represents the change in root mean square error of validation (RMSEV) and the dotted line represents the change in coefficient of determination (R^2). B: reference vs predicted casein (%) of the validation set using ENR. The solid line represents the least-square fit of the scatter plot and the dot line presents ideal prediction target (x = y).

Optimized casein content prediction models

Table 5-2 shows the optimized PLSR and ENR calibration models for casein and CN/CP quantifications with the corresponding optimization factors. For the casein content predictions, different preprocessing and regression techniques yielded similar prediction powers. PLSR and ENR showed similar model performance when predicting casein, which agreed with other studies when compared PLSR with ENR (Giglio and Brown, 2018). Unlike NIR spectral measurement which are prone to uninformative wavelength intensities, FFFS measures specific fluorophores in food matrix. Even though ENR can conduct variable selection due to specific tryptophan emission spectra, it did not improve the model performance compare to PLSR. According to Williams and Norris (2001), a R^2 above 0.95 indicates a reliable model prediction power in food analysis. The low RMSEV for the casein predictions also indicated that the model carries potential as a rapid casein quantification method. The optimized models need to be tested externally with unknown samples prior to estimate their final model performance.

Table 5-2 Optimized PLSR and ENR calibration model performance of casein and casein-to-crude-protein ratio (CN/CP) quantification

		PLSR			ENR			
	Preprocessing	R^2	RMSEV (%)	Nlv	R^2	RMSEV （%）	Alpha	Lambda
Casein (%)	SG-S	0.95	0.18	2	0.96	0.18	0.90	0.088
	SG-1st	0.97	0.14	2	0.97	0.17	0.30	0.33
CN/CP	SG-S	0.93	0.019	2	0.61	0.045	0	4.52
	SG-1st	0.93	0.018	2	0.97	0.013	0.5	0.64

SG-S = Savitzky-Golay smoothing; SG-1st = Savitzky-Golay 1st derivative; R^2 = Coefficient of determinations; RMSEV = Root mean square error of validation; Nlv = number of latent variables.

Optimized CN/CP prediction models

For CN/CP prediction, PLSR models yielded R^2 of 0.93 with both SG-S and SG-1st preprocessing techniques. The model optimizations have confirmed the validity of the calibration and optimization step, but the models need to be externally tested again to confirm its true prediction power. PLSR models are prone to overfittings, meaning that the model may perform well during the calibration and optimization step but fail to predict unknown samples (Wold et al., 2001). In the ENR predictions for CN/CP, model established with SG-S preprocessing failed the prediction with R^2 of 0.61. The RMSEV of this model was 0.045. Connecting to the previously reported CN/CP variation from 0.75 to 0.81 in pasteurized skim milk, the error range (± 0.045) from this model exceeded the natural variation of CN/CP in pasteurized skim milk. Therefore, the estimation would not have any practical use. However, with preprocessing of SG-1st, the ENR model performed well on CN/CP prediction with RMSEV of 0.013 and R^2 of 0.97, showing potential to predict unknown sample's CN/CP.

External testing of casein and CN/CP quantifications

Casein content prediction models

The externally tested model performance for casein and CN/CP quantification in UF, skim, and raw milk testing samples (Table5-3). For casein predictions, PLSR and ENR yielded similar test results with RMSEP ranged from 0.12 to 0.13% and R^2 of 0.91. To further analyze the casein quantification error, the RPD of the models ranged from 3.2 to 3.4. According to Willams and Norris (2001), RPD greater than 3 shows very good prediction power for food analysis purposes. The RPE (ranged from 4.9 to 5.1%) illustrate the relative error of the predictions in context of the average casein content (2.51%) of the test set, showing that the model estimated casein content is within 4.9 to 5.1% of the reference method. The test predictions were plotted in Figure 5-5, and the slope of least-square fit between the reference versus prediction is not significantly different from slope of 1, showing no prediction bias in casein content estimations ($P < 0.05$).

Hewavitharana (1997) first reported a Fourier-transform infrared (FTIR)-based method for casein quantification on raw milk samples. The method was developed using multivariate statistical models and validated with a set of 20 raw milk samples. The measurement range of the method was from 2.71 to 3.62% casein in raw milk, and the error of the method was reported from 0.08 to 0.1%. A follow-up study was done by Luginbühl (2002) using standard milk samples with casein range from 1.8 to 4.5%. The increased ranged also decreased the measurement error to 0.046-0.08 %. It appeared that the increase of casein calibration range and high sample homogeneity resulted in accuracy improvement in FTIR-based measurements of casein. A mid infrared (MIR)-based method was published by McDermott et al. (2016) aiming for milk casein and free amino acid quantifications. The method was designed to capture casein content variation from different genetic breeds. Though the range of the casein measurement was not reported, the error from the study was 0.48%, almost 10 times higher than the FTIR method. Though varied on individual sample predictions, the overall FFFS-based prediction results from PLSR and ENR are comparable to the existing literature. Test samples from UF, skim, and raw milk improved the robustness of the method measuring casein contents from various source of casein.

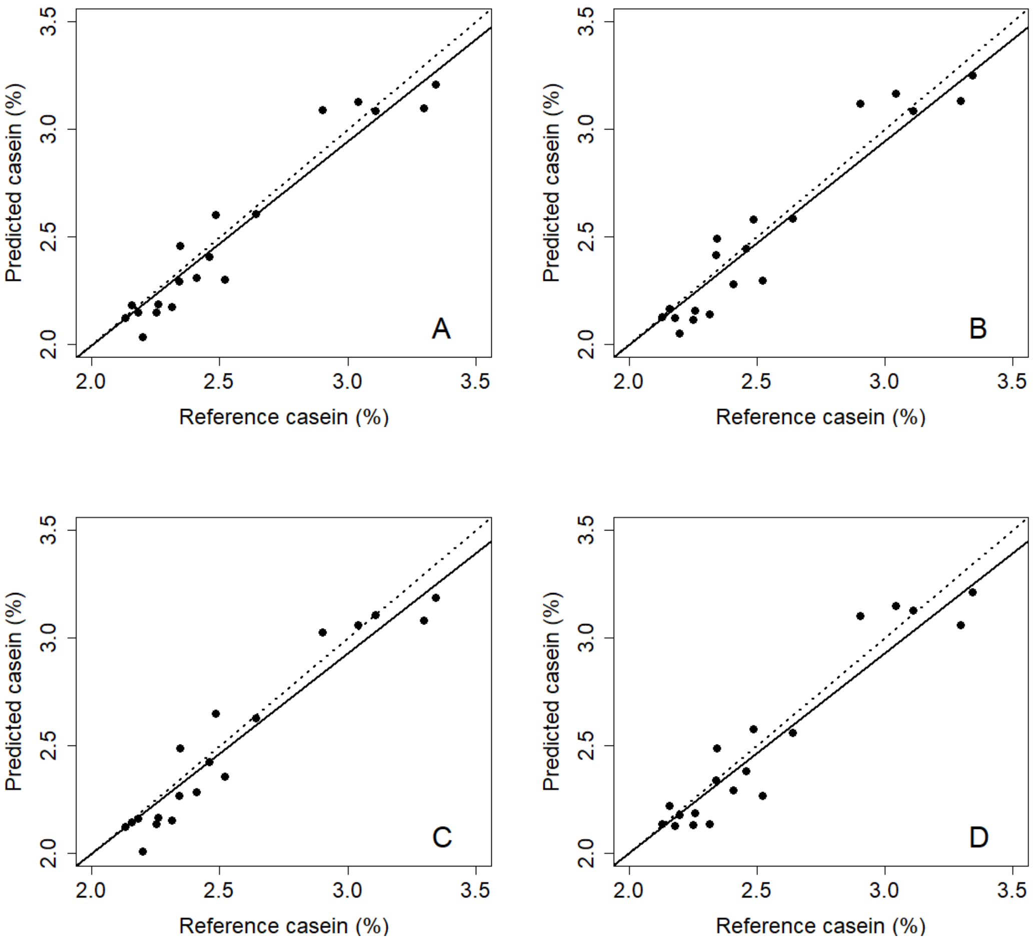


Figure 5-5 A: test model of reference versus predicted casein (%) using Savitzky-Golay smoothed (SG-S) spectra and partial least square regression (PLSR):, and B: test model of reference versus predicted casein (%) using SG-S spectra and elastic net regression (ENR), C: test model of reference versus predicted casein (%) using Savitzky-Golay first derivative (SG-1st) spectra and PLSR, and D: test model of reference versus predicted casein (%) using SG-1st and ENR. The solid line represents the least-square fit of the scatter plot and the dot line presents ideal prediction target (x = y).

CN/CP prediction models

In Table 5-3, the CN/CP test results showed similar RMSEP ranged from 0.024 to 0.027. The RMSEP are found to be higher than those yielded from the optimized models (0.013-0.019). The increase in prediction error showed that the optimized model suffered from overfitting, and the model failed to predict CN/CP in unknown milk samples (Geladi, 2002). Considering the obtained RMSEP and RPD, the model did not have any practical use in food analysis because the RPD equals to or less than 1.5. Even though the RPE of the model predictions was less than 5% from the reference values, considering the natural variation of the CN/CP (0.75-0.81) in pasteurized skim milk, the prediction values will result in large variations for CN/CP estimations. In Figure 5-6, reference versus predicted CN/CP was plotted for PLSR and ENR models. The scatters found from 0.72 to 0.80 showed that the prediction cannot yield accurate predictions for CN/CP. Therefore, with the current prediction model, it is still challenging to estimate CN/CP in UF, skim, and raw milk samples.

Table 5-3 Final PLSR and ENR test model performance of casein and casein-to-crude-protein ratio (CN/CP) quantification

		PLSR				ENR			
	Preprocessing	R^2	RMSEP (%)	RPD	RPE（%）	R^2	RMSEP (%)	RPD	RPE（%）
Casein (%)	SG-S	0.91	0.12	3.4	4.9	0.91	0.13	3.3	5.0
	SG-1st	0.91	0.13	3.2	5.1	0.91	0.13	3.3	5.0
CN/CP	SG-S	0.60	0.024	1.5	3.0	0.56	0.026	1.4	3.3
	SG-1st	0.58	0.024	1.5	3.1	0.50	0.029	1.2	3.8

SG-S = Savitzky-Golay smoothing; SG-1st = Savitzky-Golay 1st derivative; R^2 = Coefficient of determinations; RMSEP = Root mean square error of prediction; RPD = Residual prediction deviation; RPE = Relative prediction error.

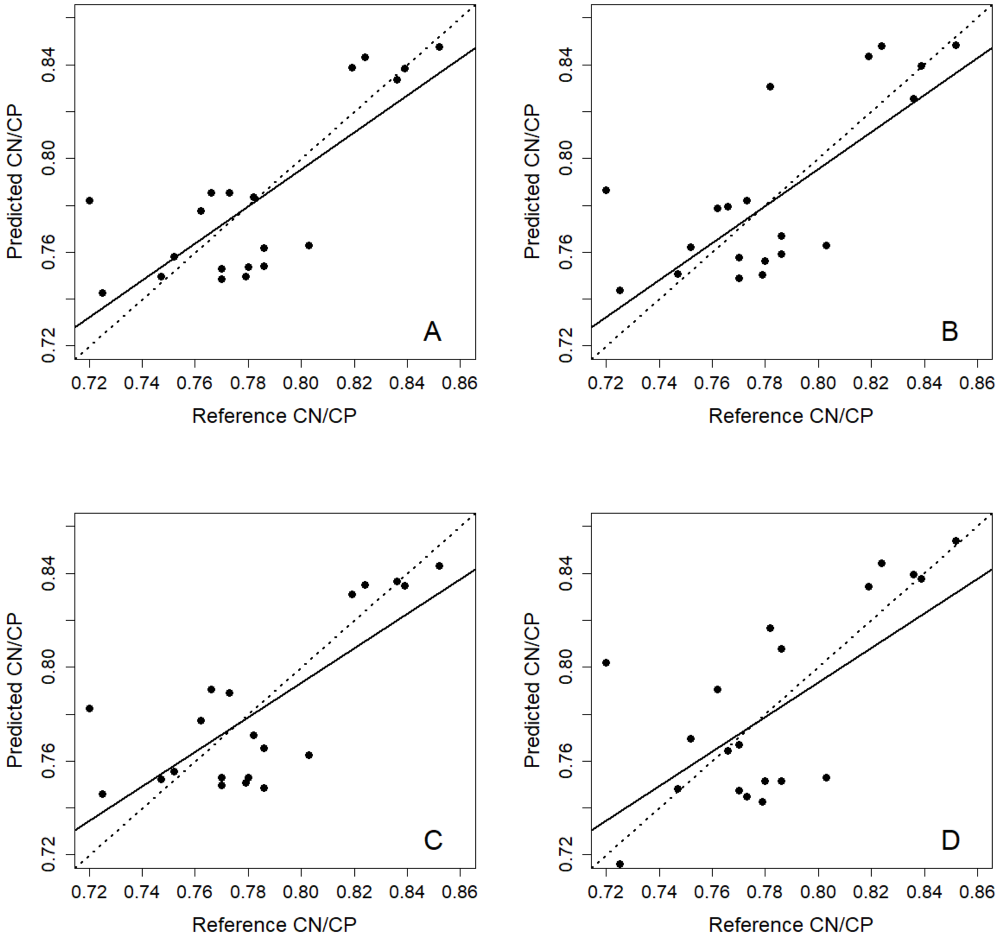


Figure 5-6 A: test model of reference versus predicted casein-to-crude-protein ratio (CN/CP) using Savitzky-Golay smoothed (SG-S) spectra and partial least square regression (PLSR):, and B: test model of reference versus predicted CN/CP using SG-S spectra and spectra and elastic net regression (ENR), C: test model of reference versus predicted CN/CP using Savitzky-Golay first derivative (SG-1st) spectra and PLSR, and D: test model of reference versus predicted CN/CP using SG-1st and ENR. The solid line represents the least-square fit of the scatter plot and the dot line presents ideal prediction target (x = y).

Conclusions

Natural variation in casein and CN/CP can impact cheese yield and quality. In this study, a FFFS-based method to measure casein and CN/CP in raw, skim, and UF milk was developed and validated. Using PLSR and ENR with external validations, the prediction models quantified casein in raw, skim, and UF milk with a RMSEP of 0.12%, an RPD of 3.4, and an RPE of 4.9% compared to the reference method. The FFFS-based method provides practical prediction power to serve the dairy industry as a rapid tool to measure casein content in fluid milk. The CN/CP quantification showed to have an RMSEP of 0.024, an RPD of 1.5, and an RPE of 3%. Though with less prediction power, the CN/CP prediction can provide preliminary indications of CN/CP variation in fluid milk. The FFFS-based method combined with chemometrics can be potentially adapted to portable fluorescence devices and implemented to dairy farms and processing plants.

References

Andersen, C.M., and G. Mortensen. 2008. Fluorescence spectroscopy: a rapid tool for analyzing dairy products. J. Arg. Food Chem. 56:720–729.

AOAC. 2016. Official methods of analysis (20th ed.). Association of Official Analytical Chemists. Gaithersburg, MD, USA: AOAC International.

Babu, K.S., and J.K. Amamcharla. 2018. Application of front-face fluorescence spectroscopy as a tool for monitoring changes in milk protein concentrate powders during storage. J. Dairy Sci. 101:10844–10859.

Barbano, D.M., and J.W. Sherbon. 1984. Cheddar cheese yields in new york. J. Dairy Sci. 67:1873–1883.

Barbano, D.M., and M.E. Dellavalle. 1987. Rapid method for determination of milk casein content by infrared analysis. J. Dairy Sci. 70:1524–1528.

Bevilacqua, M., R. Bro, F. Marini, Å. Rinnan, M.A. Rasmussen, and T. Skov. 2017. Recent chemometrics advances for foodomics. TrAC – Trend. Anal. Chem. 96:42–51.

Bonfatti, V., L. Grigoletto, A. Cecchinato, L. Gallo, and P. Carnier. 2008. Validation of a new reversed-phase high-performance liquid chromatography method for separation and quantification of bovine milk protein genetic variants. J.Chromatogr. A. 1195:101–106.

Bouveresse, E., C. Hartmann, D.L. Massart, I.R. Last, and K.A. Prebble. 1996. Standardization of near-infrared spectrometric instruments. Anal. Chem. 68:982–990.

Birlouez-Aragon, I., P. Sabat, and N. Gouti. 2002. A new method for discriminating milk heat treatment. Int. Dairy J. 12:59–67.

Brown, C.D., L. Vega-Montoto, and P.D. Wentzell. 2000. Derivative preprocessing and optimal corrections for baseline drift in multivariate calibration. Appl. Spectrosc., AS 54:1055–1068.

Chen, B., M.J. Lewis, and A.S. Grandison. 2014. Effect of seasonal variation on the composition and properties of raw milk destined for processing in the UK. Food Chem. 158:216–223.

Cipolat-Gotet, C., A. Cecchinato, M. De Marchi, and G. Bittante. 2013. Factors affecting variation of different measures of cheese yield and milk nutrient recovery from an individual model cheese-manufacturing process. J. Dairy Sci. 96:7952–7965.

Dimenna, G.P., and H.J. Segall. 1981. High-performance gel-permeation chromatography of bovine skim milk proteins. J. Liq. Chromatogr. 4:639–649.

Emmons, D.B., and H.W. Modler. 2010. Invited review: A commentary on predictive cheese yield formulas1. J. Dairy Sci. 93:5517–5537.

Filzmoser, P., M. Gschwandtner, and V. Todorov. 2012. Review of sparse methods in regression and classification with application to chemometrics. J. Chemom. 26:42–51.

Foss Electric. 1997. Application note No. 102. Calibration for casein in cow milk. MilkoCcan FT 120. Denmark.

Fox, P. F., P. L. McSweeney, and L. H. Paul. 1998. Dairy chemistry and biochemistry. London: Blackie Academic & Professional.

Friedman, J., T. Hastie, and R. Tibshirani. 2010. Regularization paths for generalized linear models via coordinate descent. J. Stat. Softw. 33:1–22.

Geladi, P. 2002. Some recent trends in the calibration literature. Chemom. Intell. Lab. Syst. 60:211–224.

Giglio, C., and S.D. Brown. 2018. Using elastic net regression to perform spectrally relevant variable selection. J. Chemom. 32:e3034.

Herbert, S., A. Riaublanc, B. Bouchet, D.J. Gallant, and E. Dufour. 1999. Fluorescence spectroscopy investigation of acid-or rennet-induced coagulation of milk. J. Dairy Sci. 82:2056–2062.

Hewavitharana, A.K. 1997. Fourier transform infrared spectrometric method for the rapid determination of casein in raw milk. Analyst. 122: 701-704.

Kulmyrzaev, A.A., D. Levieux, and É. Dufour. 2005. Front-face fluorescence spectroscopy allows the characterization of mild heat treatments applied to milk. Relations with the denaturation of milk proteins. J. Agric. Food Chem. 53:502–507.

Kumar, P., N. Sharma, R. Ranjan, S. Kumar, Z.F. Bhat, and D.K. Jeong. 2013. Perspective of membrane technology in dairy industry: a review. Asian-Australas. J. Anim. Sci. 26:1347–1358.

Lin, Y., J.A. O'Mahony, A.L. Kelly, and T.P. Guinee. 2017. Seasonal variation in the composition and processing characteristics of herd milk with varying proportions of milk from spring-calving and autumn-calving cows. J. Dairy Res. 84:444–452.

Liu, Y., W. Cai, and X. Shao. 2016. Linear model correction: A method for transferring a near-infrared multivariate calibration model without standard samples. Spectrochim. Acta A. 169:197–201.

Lucey, J., and J. Kelly. 1994. Cheese yield. Int. J. Dairy Technol. 47:1–14.

Luginbühl, W. 2002. Evaluation of designed calibration samples for casein calibration in fourier transform infrared analysis of milk. LWT - Food Sci. Technol. 35:554–558.

Lundén, A., M. Nilsson, and L. Janson. 1997. Marked Effect of β-Lactoglobulin Polymorphism on the Ratio of Casein to Total Protein in Milk. J. Dairy Sci. 80:2996–3005.

Lynch, J.M., D.M. Barbano, and J.R. Fleming. 1998. Indirect and direct determination of the casein content of milk by Kjeldahl nitrogen analysis: collaborative study. J. AOAC Int. 81:763–774.

Ma, Y.B., I. Birlouez-Aragon, and J.K. Amamcharla. 2019. Development and validation of a front-face fluorescence spectroscopy-based method to determine casein in raw milk. Int. Dairy J. 93:81-84.

McMahon, D.J., H. Du, W.R. McManus, and K.M. Larsen. 2009. Microstructural changes in casein supramolecules during acidification of skim milk. J. Dairy Sci. 92:5854–5867.

McDermott, A., G. Visentin, M. De Marchi, D.P. Berry, M.A. Fenelon, P.M. O’Connor, O.A. Kenny, and S. McParland. 2016. Prediction of individual milk proteins including free amino acids in bovine milk using mid-infrared spectroscopy and their correlations with milk processing characteristics. J. Dairy Sci. 99:3171–3182.

Mevik, B.-H., and R. Wehrens. 2007. The pls Package: principal component and partial least squares regression in R. J. Stat. Softw. 018.

R Core Team. 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Recio, I., and C. Olieman. 1996. Determination of denatured serum proteins in the casein fraction of heat-treated milk by capillary zone electrophoresis. Electrophoresis. 17:1228–1233.

Rowland, S.J. 1938. 176. The determination of the nitrogen distribution in milk. J. Dairy Res. 9:42–46.

Savitzky, A., and M.J.E. Golay. 1964. Smoothing and differentiation of data by simplified least squares procedures. Anal. Chem. 36:1627–1639.

Schaar, J., B. Hansson, and H.-E. Pettersson. 1985. Effects of genetic variants of κ-casein and β-lactoglobulin on cheesemaking. J. Dairy Res. 52:429–437.

Ur-Rehman, S., B. Kopesky, S. Backinoff, T. P. Doelman, and C. White. 2017. U.S. Pat. No. 15/446,032.

van der Ven, C., H. Gruppen, D.B.A. de Bont, and A.G.J. Voragen. 2001. Reversed phase and size exclusion chromatography of milk protein hydrolysates: relation between elution from reversed phase column and apparent molecular weight distribution. Int. Dairy J. 11:83–92.

Williams, P. C. and K. Norris. 2001. Near-infrared technology in the agricultural and food industry. St. Paul, MN: American Association of Cereal Chemists, Inc.

Walstra, P. and R. Jenness. 1984. Dairy chemistry & physics. John Wiley & Sons.

Wold, S., M. Sjöström, and L. Eriksson. 2001. PLS-regression: a basic tool of chemometrics. Chemom. Intell. Lab. Syst. 58:109–130.

Zhang, H., and L.E. Metzger. 2011. Noncasein nitrogen analysis of ultrafiltration and microfiltration retentate. J. Dairy Sci. 94:2118–2125.

Chapter 6 - Front-face fluorescence spectroscopy combined with chemometrics to detect proteinaceous matter in dairy ultrafiltration permeate[3]

Abstract

Protein leak into permeate stream can occur during ultrafiltration (UF) of milk and whey and can result in financial losses. While manufacturers can measure protein content in the finished permeate powders, there is no rapid monitoring tool during UF to identify protein leak. This study applied front-face fluorescence spectroscopy (FFFS) and chemometrics to identify the fluorophore of interest associated with the protein leak, develop predictive models to quantify true protein content, and classify the source of protein leak in permeate streams. Crude protein (CP), non-protein nitrogen, true protein (TP), tryptone-equivalent peptide, α-lactalbumin (α-LA), and β-lactoglobulin (β-LG) contents were measured for 33 lots of whey permeate and 29 lots of milk permeate from commercial manufacturers. A bimodal distribution of true-protein-to-crude-protein ratio was observed from the density plot, revealing protein leak occurrence in some whey and milk permeate. Whey permeate contained more tryptone-equivalent peptide than milk permeate, while α-LA and β-LG were only detected in 59% milk permeate. The source of protein leak was therefore identified and labeled for predictive model development. Representative permeate with high and low TP contents were analyzed using excitation-emission matrix (EEM) and tryptophan was identified as the fluorophore of interest for protein leak detection. Tryptophan excitation spectra were collected for predictive model development using partial

[3] Submitted for publication: Journal of Dairy Science

least square regression to quantify the true protein content in the permeate. With external validations, a useful model for quality control purposes was developed with a root mean square error of prediction of 0.22% (dry-basis) and a residual prediction deviation of 2.8. Moreover, classification models were developed using partial least square discriminant analysis, and the classification methods can detect high TP level, high tryptone-equivalent peptide level, and presence of α-LA or β-LG with 83.3%, 84.8%, and 98.5% cross-validated accuracy, respectively. This method showed that FFFS and chemometrics can rapidly detect protein leak and identify the source of protein leak in UF permeate. Implementation of this method in UF processing plants can reduce financial loss from protein leak and maintain high-quality permeate production.

Introduction

Ultrafiltration (UF) is commonly used to fractionate proteins from skim milk and cheese whey, resulting in protein-rich retentate and lactose-rich permeate. The protein-rich retentate is further processed into milk protein concentrate/isolate and whey protein concentrate/isolate in powder forms. The lactose-rich permeate of milk or whey is then processed into milk permeate powder (MPP) or deproteinized whey (DPW) powder, respectively. Since diafiltration is often performed during UF, the resulting permeate is diluted with various amounts of water and consequently the permeate composition can vary largely among different manufacturers. Therefore, the composition of permeates is normally reported on a dry basis after the liquid permeate is concentrated, crystalized, and spray dried. According to the American Dairy Products Institute (ADPI), permeate powder contains at least 76% lactose and at most 7% crude protein (CP), among other components. The CP may include non-protein nitrogen (NPN) and some fractions of proteinaceous matter such as peptides and proteins from milk or whey (Walstra et al., 2006).

A fraction of proteinaceous matter can pass through the membrane into the permeate during UF. Because protein is the most profitable component in the UF processing, leaking of protein into the permeate results in financial losses. In addition, the presence of protein in permeate powder can accelerate loss of quality during storage (Sithole et al., 2005). However, detection of CP in the permeate powders only occurs after processing several batches at industrial scale resulting in a considerable lag time between the occurrence of protein leak and corrective action needed to prevent any protein leak. Currently, there are limited at-line or on-line technologies to monitor the quality of the permeate stream and detect abnormal levels of proteinaceous matter during UF. Therefore, an accurate and rapid method is needed to detect the protein leak during UF.

Front-face fluorescence spectroscopy (FFFS) is known for its high sensitivity and ability to analyze turbid samples. Milk contains intrinsic fluorescence compounds such as tryptophan, nicotinamide adenine dinucleotide (NADH), vitamin A, and Maillard browning products which are identified as fluorophores of interest to characterize dairy products (Shaikh and O'Donnell, 2017). Multivariate statistical analyses were used to extract quantitative, qualitative, or structural information from fluorescence spectra (Karoui et al., 2011). FFFS has been used for protein detection in dairy products. Kulmyrzaev et al (2005) has used FFFS to characterize thermal effects on different milk proteins during pasteurization. To expand the applications of FFFS, supervised chemometric models can provide quantification and classification of non-fat dry milk, milk protein concentrate, and crude synthesis (Liu and Metzger, 2007; Babu and Amamcharla, 2018; Fagan et al., 2011).

Moreover, fluorescence spectroscopy can detect conformational differences in proteins (Vivian and Callis, 2001). The band shape, maximum intensity, and the fluorescence lifetime are sensitive to polypeptide structures and chemical environment (Royer, 2006). In some instances, UF membrane failure can lead to irreversible changes to the membrane porous size, which can lead to full protein (α-LA and β-LG) leak through the membrane. UF feeds (skim milk and whey) are also prone to proteolysis coming from enzyme residue and microbial growth, leading to various sizes of protein hydrolysate. The protein hydrolysates are small in molecular size which allows them to pass through properly functioned UF membranes. FFFS spectra may provide information on the source of protein leak in permeate from either full protein or peptides due to its specificity to protein conformations.

The objectives of this study were to identify the fluorophore of interest associated with the protein leak and develop predictive models to quantify true protein content and classify the source of protein leak in permeate streams.

Materials and methods

Experimental approach

Milk and whey permeate powders were randomly collected from 2 commercial manufacturers (37 lots of DPW and 29 lots of MPP) and were analyzed for crude protein (CP), non-protein nitrogen (NPN), and true protein (TP). Based on the true-protein-to-crude protein ratio (TP/CP), permeates were classified as either high or low-TP permeate. In addition, tryptone-equivalent peptides, α-LA, and β-LG were also carried out to further understand the TP fraction of the permeates and were used to identify the source of protein leak. Non-target FFFS analysis of permeates discovered the available fluorophores and identified the fluorophore of interest in representative high and low TP permeates. Chemometric models were developed based on the permeate classes and FFFS spectra of the target fluorophore. Detail experimental procedures are included in the following sections.

Permeate samples

Milk and whey permeate powders were randomly collected from 2 commercial manufacturers (37 lots of DPW and 29 lots of MPP). The permeate powder was reconstituted to 5% solution (w/w) at 25°C for 1 hour. The 5% concentration was selected to simulate the permeate stream coming from the UF process and were used for all subsequent analyses unless specifically notified. Solutions made from MPP were referred as milk permeate and solutions made from DPW were referred as whey permeate. Collectively, permeates were used to indicate both milk and whey permeates.

True-protein-to-crude-protein ratio (TP/CP ratio)

CP and NPN of the permeates were analyzed by the Kjeldahl nitrogen method (AOAC Methods 991.20 and 991.21) in duplicates with a nitrogen conversion factor of 6.38. TP was calculated as the difference between CP and NPN. TP/CP as calculated using the TP divided by the CP. The proportion of the TP in CP determines the protein leak in permeates. A density plot of TP/CP was produced to identify protein leak occurrence and classify permeates as high- and low-true-protein (HTP and LTP), respectively.

True protein fraction characterizations

Tryptone-equivalent peptide quantification

Tryptone-equivalent peptide content in permeates was quantified by a fluorescence-based method developed by Udenfriend et al. (1972) with some modifications. In a microplate, 150 µL permeates and 50 µL fluorescamine (Sigma Aldrich, St. Louis, USA) reagent solution (10.8 mM in acetone) were added to each well. The microplate was shaken for 1 min in dark prior to the fluorescence measurement using a spectrofluorometer (BioTek, Winooski, USA) with a 355 nm excitation filter and 460 emission filter at 25°C. Tryptone-equivalent peptide content was quantified using a series of tryptone standards (Sigma Aldrich, St. Louis, USA) in duplicates and expressed as tryptone equivalent.

α-LA and β-LG analysis

α-LA and β-LG were quantified using an HPLC-based method developed by Bonfatti et al. (2008) with modifications. Briefly, permeate powders were dissolved in 6 M Guanidine hydrochloride (5% w/w) at 25°C for 1 hour. The solution was filtered through a membrane filter (Millipore Millex®-FH, Polyethylene, 0.22μm pore size) and 8 μL of the filtered sample was injected into a HPLC system equipped with a diode-array detector (DAD; HP 1050 series, Palo Alto, CA)). Separation was achieved by using a silica-based packing C8 column (Zorbax 300SB-C8 RP, 3.5 μm, 300 Å, 150 × 4.6 I.D., Agilent Technologies) and a pre-column (Widepore C4 4 mm × 3.0 mm, Phenomenex). A gradient elution was performed using two solvents. Solvent A consisted of 0.1% trifluoroacetic acid (TFA) in water and solvent B was 0.1% TFA in acetonitrile. Total analysis time was 45 min beginning with a linear gradient from 33 to 35% B in 5 min, from 35 to 37% B in 4 min, from 37 to 40% B in 9 min, from 40 to 41% B in 4 min, followed by an isocratic elution at 41% B during 5.5 min, then linear gradient from 41 to 43% B in 0.5 min, from 43 to 45% B in 8 min, and return linearly to the starting condition in 1 min. The column was re-equilibrated under the starting conditions for 8 min prior to the next injection. The analysis was done at a flow rate of 0.5 ml/min and 45 °C. The detection was made at 214 nm using the DAD detector.

Pure whey protein standards (> 99%) of α-LA and β-LG were purchased from Sigma–Aldrich (St. Louis, MO, USA). Standard solutions were prepared in 6 M Guanidine hydrochloride using the same procedure described above, and α-LA and β-LG quantifications of permeate were achieved using the external standard curves.

Front-face fluorescence spectroscopy analysis

Excitation-emission-matrix of permeate

Non-targeted FFFS analysis of representative HTP and LTP whey and milk permeates was achieved using the excitation-emission matrix (EEM). To collect the EEM, 7 mL of permeates was transferred into a quartz cuvette (Starna Cells Inc., Atascadero, CA). The front-face EEM was acquired at 25°C by recording the emission (λ_{Em}) spectra from 300 to 500 nm (with 0.5 nm interval) corresponding excitation wavelengths (λ_{Ex}) ranging from 220 to 360 nm (with 10 nm interval) using a spectrofluorometer (LS-55; Perkin Elmer, Waltham, MA). The emission and excitation slit width were set at 4 nm and scan speed was set at 100 nm/min.

Excitation spectra of tryptophan

After preliminary examination of the EEM, tryptophan was selected as the fluorophore of interest. Tryptophan front-face excitation spectra (λ_{Em} at 340 nm; λ_{Ex} 200-300 nm) were collected using the spectrofluorometer at 25°C. The emission and excitation slit width was set at 5 nm and the scan speed was set at 300 nm/min. For each sample, triplicate analysis was carried out and the scans were averaged to improve signal-to-noise ratio.

Spectral preprocessing and principal component analysis

The tryptophan fluorescence excitation spectra were normalized by the standard normal variate (SNV) technique prior to chemometric analysis. This normalization process unified fluorescence intensity scale and minimize bias for the subsequent chemometric analysis (Li et al., 2015). Principal component analysis (PCA) was applied to reduce the dimensionality and redundancy in the dataset, and the PCA similarity plots were generated to provide visualization of permeate sample clustering.

Partial least square regression and discriminative analysis

Partial least squares regression (PLSR) is a supervised chemometric algorithm which has been widely used to model spectral data (Diez et al., 2008). PLSR extracts latent variables from the spectral data and maximize their covariance to the reference values. In this study, TP content was predicted using PLSR with SNV-normalized tryptophan excitation spectra as input. The original permeate dataset was partitioned randomly to a calibration set (n_{cal} = 47) and a validation set (n_{val} = 19). The calibration models were developed and validated using the leave-one-out cross validation (LOOCV). The cross-validated model was evaluated by root mean square error of cross-validation (RMSECV) as shown in Equation 1. Residual prediction deviation (RPD) was calculated by dividing the RMSECV by the standard deviation of the reference dataset to estimate model prediction power (Williams and Norris, 2001). The calibration model was also externally tested by predicting the validation set. Root mean square error of prediction (RMSEP) and RPD of the prediction confirmed the model validity on unseen samples, which can estimate the future model applicability of TP quantification (Wold et al. 2001).

$$RMSECV/P \ = \sqrt{\frac{\sum_{j=1}^{n}(y_{j-} \ \hat{y}_j)^2}{n}} \qquad [1]$$

Where y is the predicted value from cross validation (CV) or prediction (P) of the jth sample, y is the reference value of the jth sample, n is sample size.

Classification models using partial least squares discriminative analysis (PLS-DA) were developed to identify the source of protein leak. PLS-DA is a supervised classification method based on linear discriminant models (Lenhardt et al., 2015). The reference variables in PLS-DA are dummy variables indicating whether 1 sample belongs to a certain class. In this study, the reference classes used for PLS-DA included TP levels, tryptone-equivalent peptide levels, and α-

LA or β-LG presence/absence. Using the same predictor as PLSR, the classification models were cross-validated using LOOCV. Accuracy, sensitivity, and specificity were used to evaluate model performance.

Statistical analysis

Unpaired t-test was used to compare means of CP, TP, and tryptone-equivalent peptide between milk and whey permeates. T-tests, chemometric visualization and modelling were performed using RStudio (version 1.1.442; RStudio, Boston, MA) with the Caret and pls packages (Kuhn, 2008; Mevik and Wehrens, 2007)

RESULTS AND DISCUSSION

Proteinaceous matter characterization

Proteinaceous fractions of permeate powders

Even though UF has been a common practice in dairy food processing, protein analysis of permeate powders has not been widely reported. As shown in Figure 6-1A, the average NPN contents in whey and milk permeate were found to be 2.51 and 2.35% (expressed as protein-equivalent on dry basis), respectively, with no significant difference ($P > 0.05$). Smith et al. (2016) reported the NPN in UF cheese permeate to be between 2.5 to 3.6% (dry basis) using the Kjeldahl-based method. The variation in NPN is related to the cheese making process, and low-pH whey with more proteolysis can lead to increase of soluble nitrogen. Smith et al. (2016) also measured TP in the whey and milk permeates ranged from not detectable to 0.003 ± 0.003%, showing negligible TP presence in permeate powders. However, the pilot-scale permeate manufacturing process featured in that study was designed for research purposes. The minimal TP content can represent the ideal situation of the UF efficiency. Frankowski et al. (2014) collected spray-dried and liquid whey permeate from commercial manufacturers. Though some permeates had no TP detected, the reported maximum TP found in that study was 0.88% (dry basis). In this study, whey and milk permeates showed a higher TP content for whey and milk permeates with averages (±SD) of 0.69 ± 0.48 and 1.17 ± 0.51 %, respectively ($P < 0.05$). The TP contents were higher than the previously reported values because the samples were collected from the commercial production runs. Due to the large variation in TP, CP in whey and milk permeate ranged from 2.51 to 3.40% and 2.72 to 3.58%, respectively, with no significant difference ($P > 0.05$). The variation in TP content from whey and milk permeates showed that protein leak may have occurred in some of the samples considered in the study.

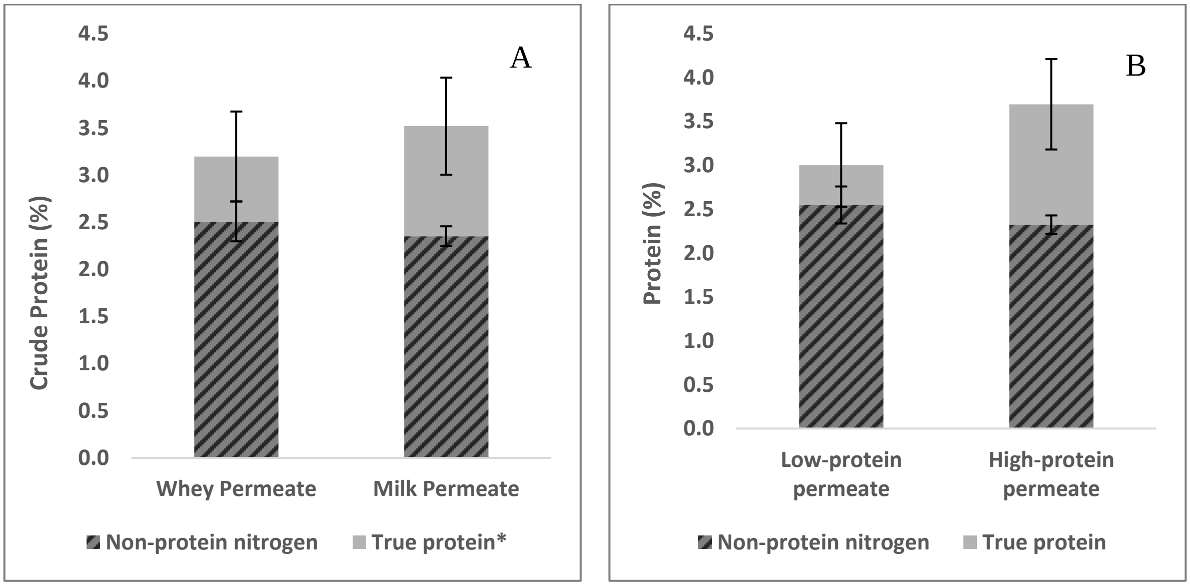


Figure 6-1 Protein fraction comparison between A: whey permeates and milk permeates, B: low-true protein permeates and high-true protein permeates. *($P < 0.05$).

Relatively a large variation in TP indicated existence of subgroups among the permeates. The TP/CP ratio was calculated to highlight TP differences while minimizing the variation in CP and shown in Figure 6-1B. Figure 6-2 shows the TP/CP ratio density plot of the permeates as a bimodal distribution with 2 distinct groups. The mean of the TP/CP ($\mu = 0.25$) was used as a cut-off to classify permeates as low-true-protein permeate (LTP, where TP/CP < 0.25, n = 34) and high-true-protein permeate (HTP, where TP/CP > 0.25, n = 32). Figure 6-1B illustrated protein fractions of LTP and HTP. The NPN fraction of LTP and HTP was similar to the prior comparison, but larger difference was observed in the TP fraction indicating the occurrence of protein leak in the HTP group. Among the LTP permeates, there is still some proteinaceous matter in the sample indicating presence of proteins or peptides. even though with a reduction of average TP content (0.46%).

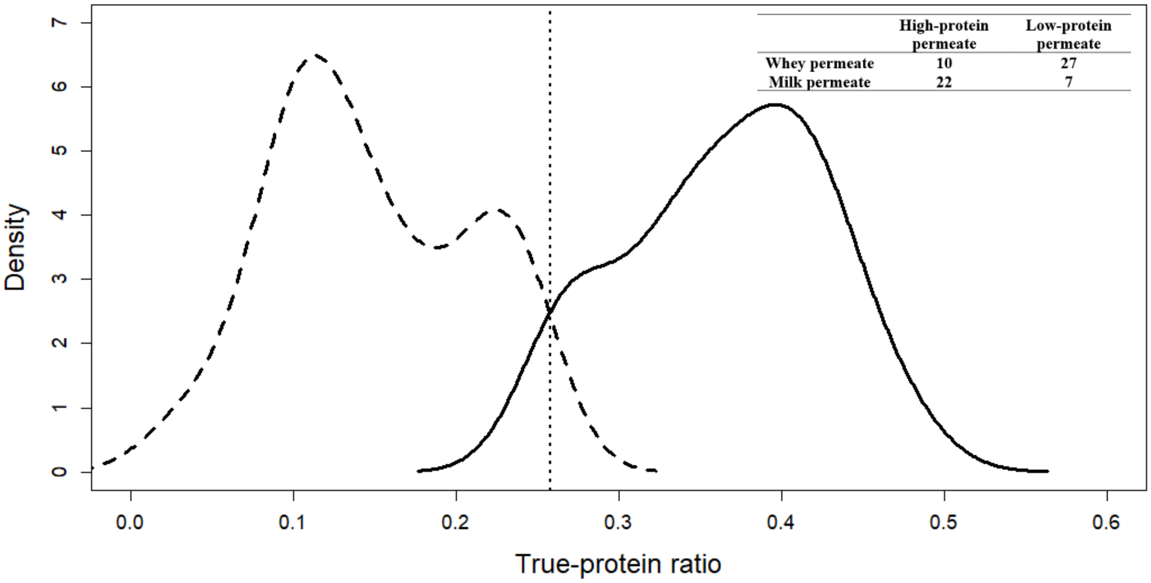


Figure 6-2 Density plot of true protein ratio (n =66). The dashed line represented the distribution of low-protein permeate (n = 34) and the solid line represented the distribution of the high-protein permeate (n = 32). The vertical dot-line represented the mean of the overall true protein ratio (μ = 0.25).

Tryptone-equivalent peptide analysis

The average peptide content of permeates using tryptone as standard is shown in Figure 6-3A. The fluorometric method utilized a primary amine reaction to attach a fluorescence dye and quantify peptides or proteins using the corresponding standard. Tryptone has been used as the standard measure peptide content and degree of protein hydrolysis in dairy foods (Minervini et al., 2003; Oh et al., 2013). Therefore, tryptone equivalence can be used as an effective measurement of peptide content in permeates. It should be noticed that the tryptone-equivalent peptides quantified in permeates were less than 10 kDa since the samples have gone through UF. The leak of proteinaceous matter as tryptone-equivalent peptides (< 10 kDa) does not indicate membrane performance failures, but it may correspond to proteolysis prior to UF. Whey permeate showed an average tryptone-equivalent peptide content (±SD) of 0.37 ± 0.04% tryptone equivalent) and milk permeate showed an average tryptone-equivalent peptide content (±SD) of 0.32 ± 0.07% tryptone equivalent; P < 0.05). The higher tryptone-equivalent peptide content in whey permeate agreed with the proteolysis process from cheese making. Residual rennet and cheese culture are both proteolytic (Fox, 1989). Especially after whey draining, timely inactivation of rennet and bacteria are needed to minimize proteolysis occurrence. On the other hand, to produce milk permeate, skim milk is directly processed as the feed for UF, so less exposure of proteolytic enzymes or bacteria will occur in UF of milk.

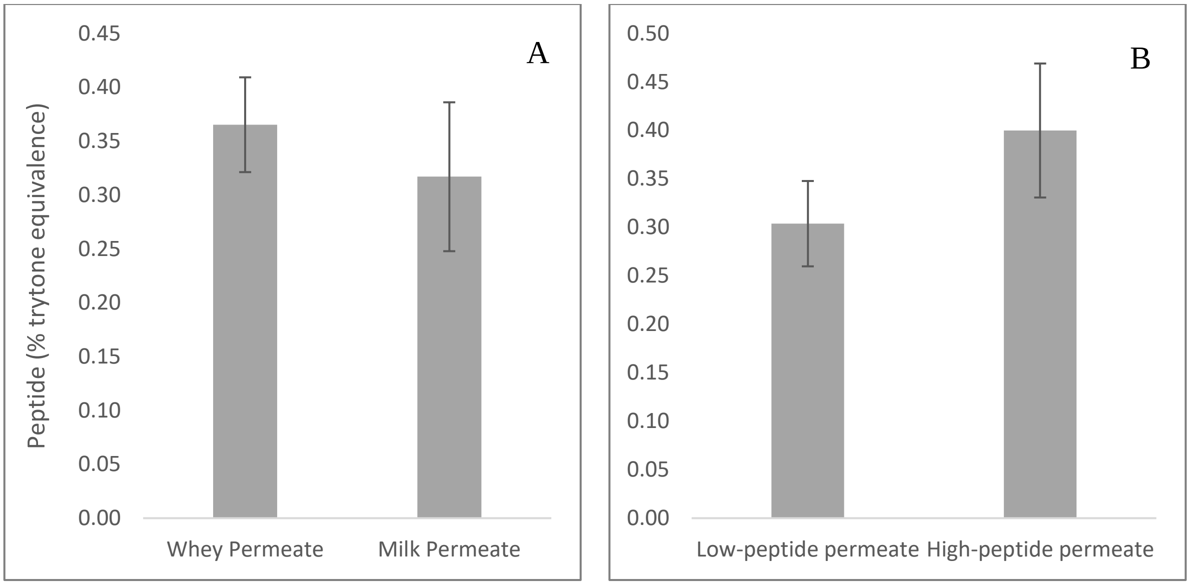


Figure 6-3 Tryptone-equivalent peptide comparison between A: whey permeates and milk permeates, B: low-peptide permeates and high-peptide permeates.

Two subgroups with high- and low-tryptone-equivalent peptide contents in Figure 6-3B. The subgrouping was achieved by plotting tryptone-equivalent peptide content on a density plot. In Figure 6-4, the distribution of the tryptone-equivalent peptide content is shown, and high- and low-tryptone-equivalent peptide permeates were classified base on the average dry-basis content of 0.35%. Subgrouping shows 25 out of 37 whey permeates were classified as high-tryptone-equivalent peptide permeate, whereas only 5 milk permeates were in the high-peptide category. On the other hand, 24 out of 29 milk permeates was classified as low-peptide permeate, which agrees with the lesser proteolysis occurred. The high- and low-peptide permeate class allows supervised chemometric model development to predict the source of protein leak.

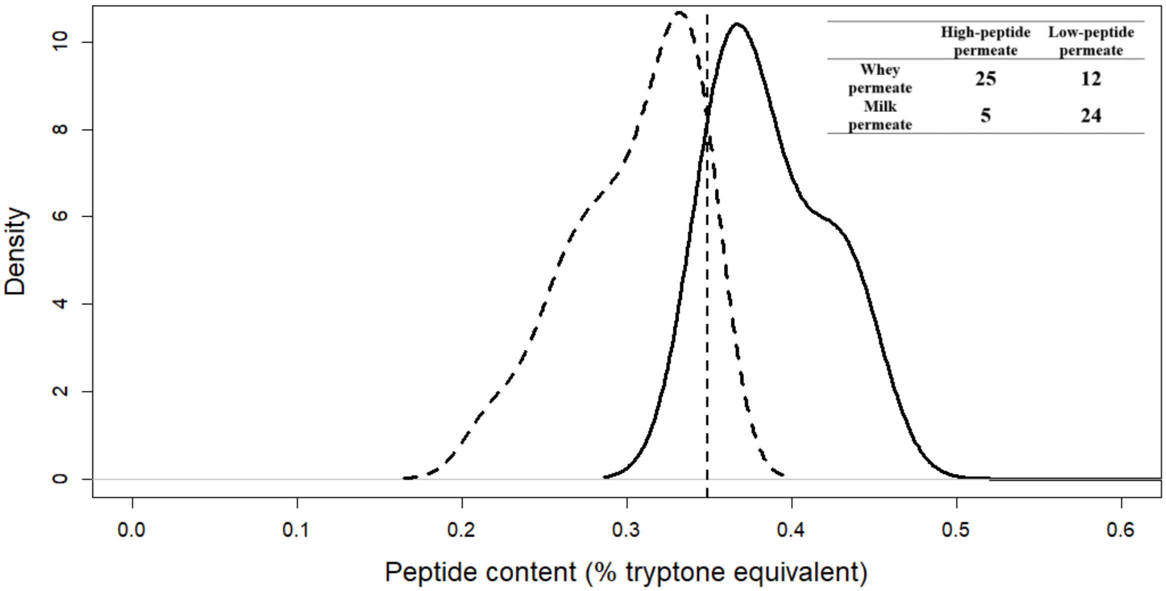


Figure 6-4 Density plot of tryptone-equivalent peptide content (n =66). The dashed line represented the distribution of low-peptide permeate (n = 31) and the solid line represented the distribution of the high-peptide permeate (n = 35). The vertical dot-line represents the mean of the overall peptide content (μ = 0.35% tryptone equivalent).

α-LA and β-LG analysis

HPLC analysis of α-LA and β-LG revealed the full protein presence in the permeates. Overall, whey permeates had no detectable α-LA or β-LG. In milk permeate, 25 and 17 lots contained detectable levels of α-LA and β-LG, respectively. The detection limits of the method by Bonfatti et al. (2008) were 0.6 µg for both whey protein fractions. On a dry basis, in Figure 6-5, more α-LA (μ = 0.27%) was found in the milk permeate than β-LG (μ = 0.04%), and the presence of full protein indicated another source of protein leak. The UF membrane contains a pore distribution with the average pore cutoff reported as 10 kDa; however, depending on the processing conditions such as transmembrane pressure, the membrane pore cutoff is altered and consequently some whey proteins can pass through. Based on Figure 6-5, more α-LA (~14kDa) was detected in the permeate than β-LG (~18kDa), agreeing with the hypothesized pore cutoff expansion for smaller protein during UF. Considering the similar true protein content between whey and milk permeates, more TP fractions of whey permeates may come from peptide rather than full whey protein, in agreement with the tryptone-equivalent peptide data. Milk permeates contained full α-LA and β-LG, which contributed to the true protein content shown in Figure 6-1. For chemometric predictions, the presence and absence of α-LA or β-LG was used as a binary class for the permeate samples.

The bimodal distribution of TP/CP, presence of peptides, α-LA, and β-LG indicated that protein leak occurred in some permeates. The sources of protein leak can vary based on either whey or milk as the UF feed material. Subsequent FFFS analysis aimed to identify the fluorophore of interests associated with true protein presence in permeates. Chemometric analysis will further model spectral data to predict true protein contents and identify sources of protein leak.

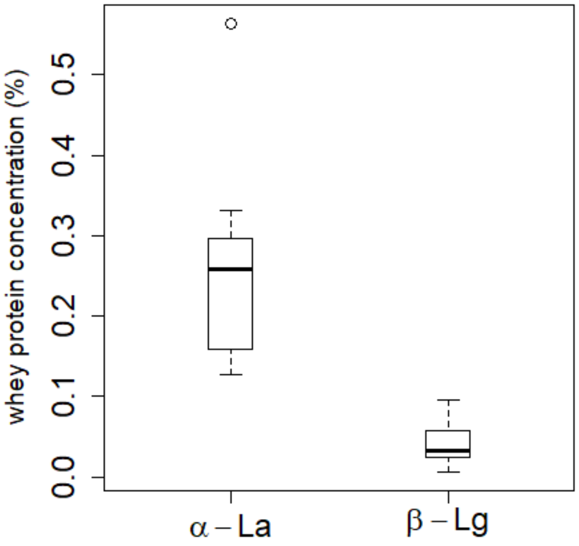


Figure 6-5 whey protein fractions: α-LA (n = 25) and β-LG (n = 17) presence in permeate samples.

Fluorescence spectra of HTP and LTP permeates

Excitation-emission-matrix of permeate stream

The EEM (Figure 6-6) characterizes representative HTP and LTP permeates of whey and milk in terms of differences and similarities in their fluorescence intensities at various excitation and emission wavelengths. There were 4 fluorescence maxima found on the EEM (Figure 6-6) of permeates corresponding to 3 fluorophores commonly found in permeates. In Figure 6-6A and B, riboflavin appeared in HTP and LTP whey permeates at fluorescence maxima λ_{Ex} 270 nm and λ_{Em} 430 nm. Andersen and Mortensen (2008) summarized common fluorophores in dairy products and riboflavin was reported to have three maxima at λ_{Ex} 270, 360, and 450 nm and λ_{Em} 430 nm. However, riboflavin showed similar peak intensity for both HTP and LTP whey permeate and consequently riboflavin was not selected as the fluorophore of interest for protein leak detection. Fluorescence intensities of the Maillard browning products are found in a range from λ_{Ex} from 310 to 350 nm and λ_{Em} from 380 to 430 nm for whey permeate (Figures 6-6A and

6-6B). Fluorescence of Maillard browning products in dairy products have been widely reported to characterize heat treatments and storage changes in dairy foods. Kulmyrzaev et al. (2005) applied FFFS for heat-induced changes of protein concentration in pasteurized milk, and Babu and Amamcharla (2018) measured Maillard browning FFFS spectra to characterize storage-induced solubility changes in milk protein concentrate. However, the fluorescence intensity from Maillard browning products can be influenced by other factors such as storage time and temperature exposure history of the permeate powders (Birlouez-Aragon et al., 2002). Therefore, Maillard browning products may not be suited as the fluorophore of interest for protein leak detection and were not considered for further chemometric analysis. A similar trend of riboflavin and Maillard browning products EEM appearance was also observed in Figures 6-6C and 6-6D for milk permeates.

There are 2 regions in the permeate EEM corresponding to the tryptophan fluorescence and labeled as T1 and T2 in Figure 6-6A. The T1 with intensity maxima at λ_{Ex} 280 nm and λ_{Em} 340 nm is commonly measured in FFFS analysis of food products (Karoui and Blecker, 2011). This tryptophan region has been reported as indicators for classifying cheese origins and storage effects on non-fat dry milk (Liu and Metzger, 2007; Karoui et al., 2006). Region T2 (Figure 6-6A) with maxima λ_{Ex} 225 nm and λ_{Em} 340 nm has been reported as the tryptophan side chain fluorescence corresponding to tryptophan-containing peptides (Becker, 2005). Whey permeate (Figure 6-6B) and milk permeate (Figure 6-6D) indicated high fluorescence intensities in both T1 and T2 regions for the HTP permeates compared to the LTP permeates. The consistent intensity difference between HTP and LTP permeates indicated that tryptophan has potential as the fluorophore of interest for protein leak detection. To enrich the spectral information collected, tryptophan excitation spectra can highlight both tryptophan peak intensities with one single scan.

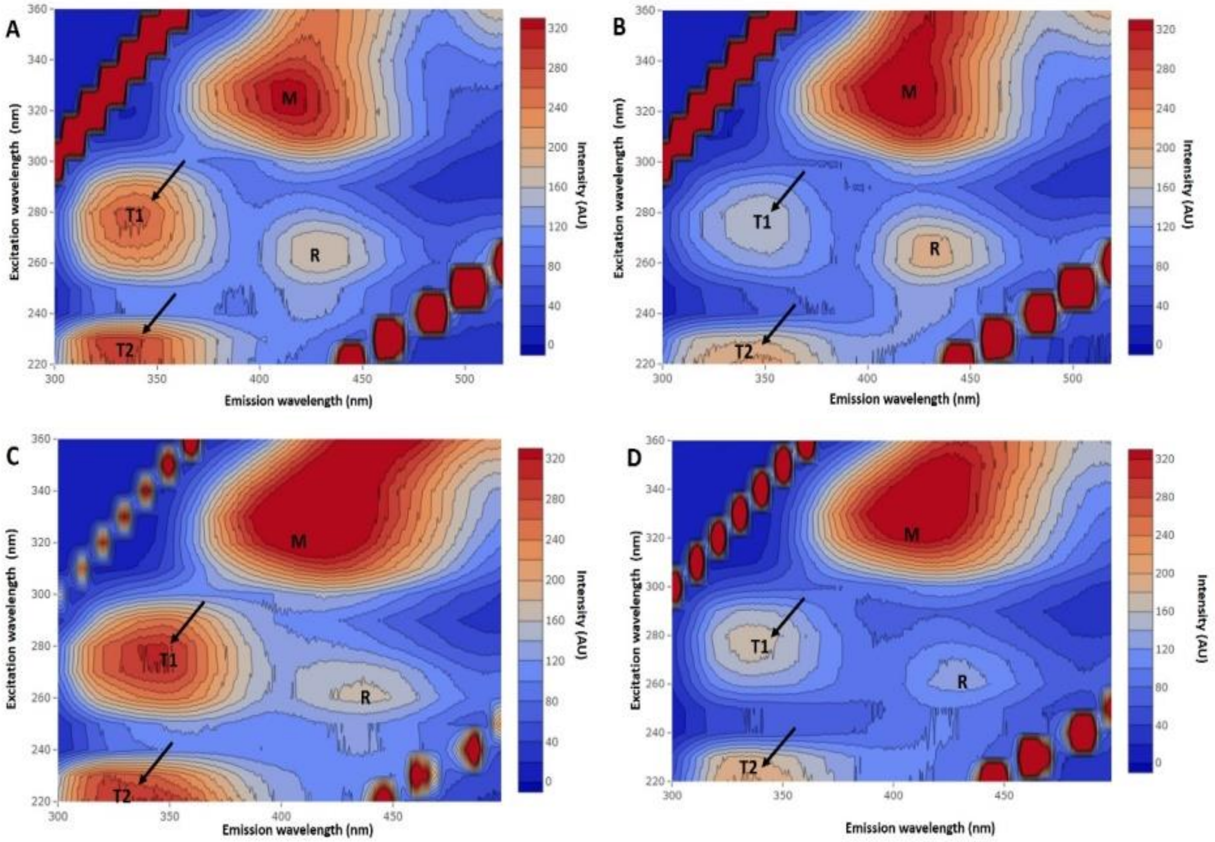


Figure 6-6 Excitation-emission matrices of A: high-true-protein whey permeate, B: low-true-protein whey permeate, C: high-true-protein milk permeate, and D: low-true-protein milk permeate. Area T1: tryptophan fluorescence from protein, area T2: tryptophan fluorescence from peptides, area M: Maillard browning products, and area R: riboflavin. Arrows point the major intensity differences among the samples.

Excitation Spectra of Tryptophan

Tryptophan excitation spectra were collected based on the intensity maxima found from Figures 6A-D. Figure 6-7A shows representative baseline-corrected excitation spectra (λ_{Em} at 340 nm) of LTP and HTP permeates characterizing the spectral variation in both tryptophan regions. The maximum fluorescence intensity was found at 227 nm and spectral variation occurred at 220-225 nm and 260-290 nm. The intensity maxima region from 260-290 nm corresponds to tryptophan-containing proteins found in dairy products (Karoui and Blecker, 2011). Backer (2005) reported that fluorescence maxima from 220-225 nm represented the tryptophan-containing amino acids and peptides. The 2 tryptophan regions agreed with the EEM and the tryptophan fluorescence spectra can characterize the protein leak in UF permeates. The spectral variation showed a similar trend (Figure 6-7B) with less variation observed in the high- and low-peptide permeates, corresponding to the tryptone-equivalent peptide difference characterized by FFFS (region T2). Therefore, chemometric analysis used tryptophan excitation spectra as inputs to classify HTP and LTP permeates, high- and low-peptide permeates, presence and absence of α-LA or β-LG in permeates and quantifying true protein content in permeates.

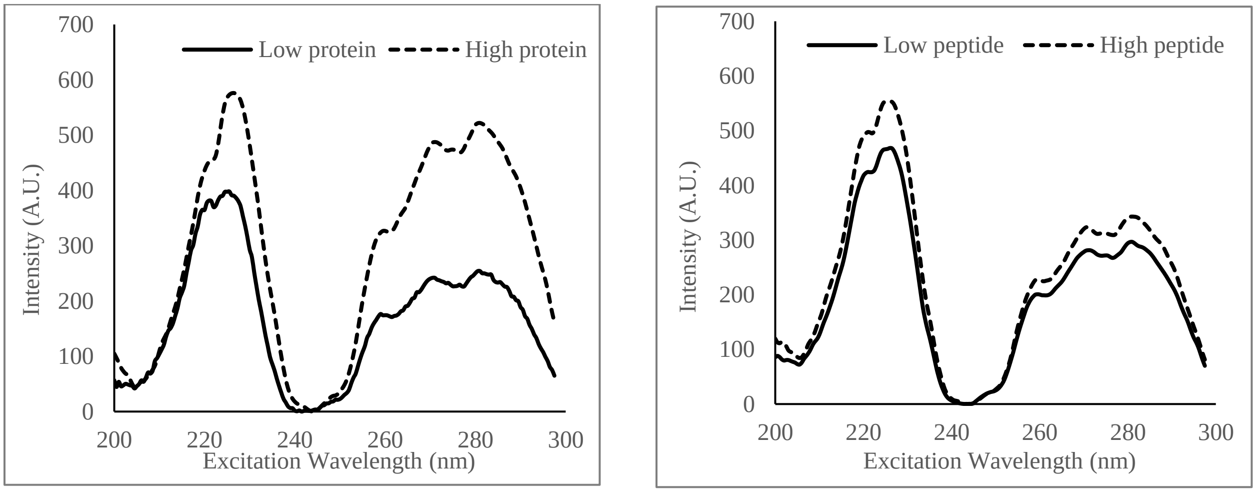


Figure 6-7 Representative tryptophan excitation fluorescence spectra of A: low-protein and high-protein permeate solution and B: low-peptide and high-peptide permeate.

Chemometric analysis

Principal component analysis

Tryptophan excitation spectra were used to extract principal components. The first, second, and third principal components (PC1, PC2, and PC3) explained 95% of the spectral variation in the tryptophan excitation spectra. In Figure 6-8A, the similarity plot using PC1 and PC2 is shown to characterize TP changes among permeate samples. Based on the differences highlighted from tryptophan peak intensities in Figure 6-7A, the TP difference among the permeates can be described by the arrow shown on Figure 6-8A. Variations in fluorescence intensity came from both protein and tryptone-equivalent peptide contents in the permeates as they contribute to TP content measured by the Kjeldahl method. PCA has shown robust dimension reduction from various studies based on tryptophan fluorescence spectra (Kulmyrzaev et al., 2005; Karoui et al., 2006). Data obtained from this study confirmed tryptophan fluorescence as a reliable source to quantify dairy proteins. The overall TP increase trend from PCA also indicated potential supervised regression prediction for the TP contents in permeates.

According to Figure 6-6-8B, the low and high-peptide permeates classifications are used to label the similarity map. After preliminary visual comparison, PC1 and PC3 were used to best-visualize the different types of permeate based on tryptone-equivalent peptide contents. From Figure 6-8B, it is observed that the high and low-peptide permeates are not well separated using PCA while some low-peptide permeates are found in the high-peptide cluster. Connecting to Figure 6-7B, the peptide fluorescence intensity differences are not reflected as largely as the TP difference due to the tryptophan-containing protein’s signal interference with the peptide. However, fluorescence spectra are sensitive enough to differentiate peptide and whey protein compositions from the permeate source. The clusters observed from the similarity plot indicated the feasibility of classification models to predict low or high-peptide permeate for unknown sources.

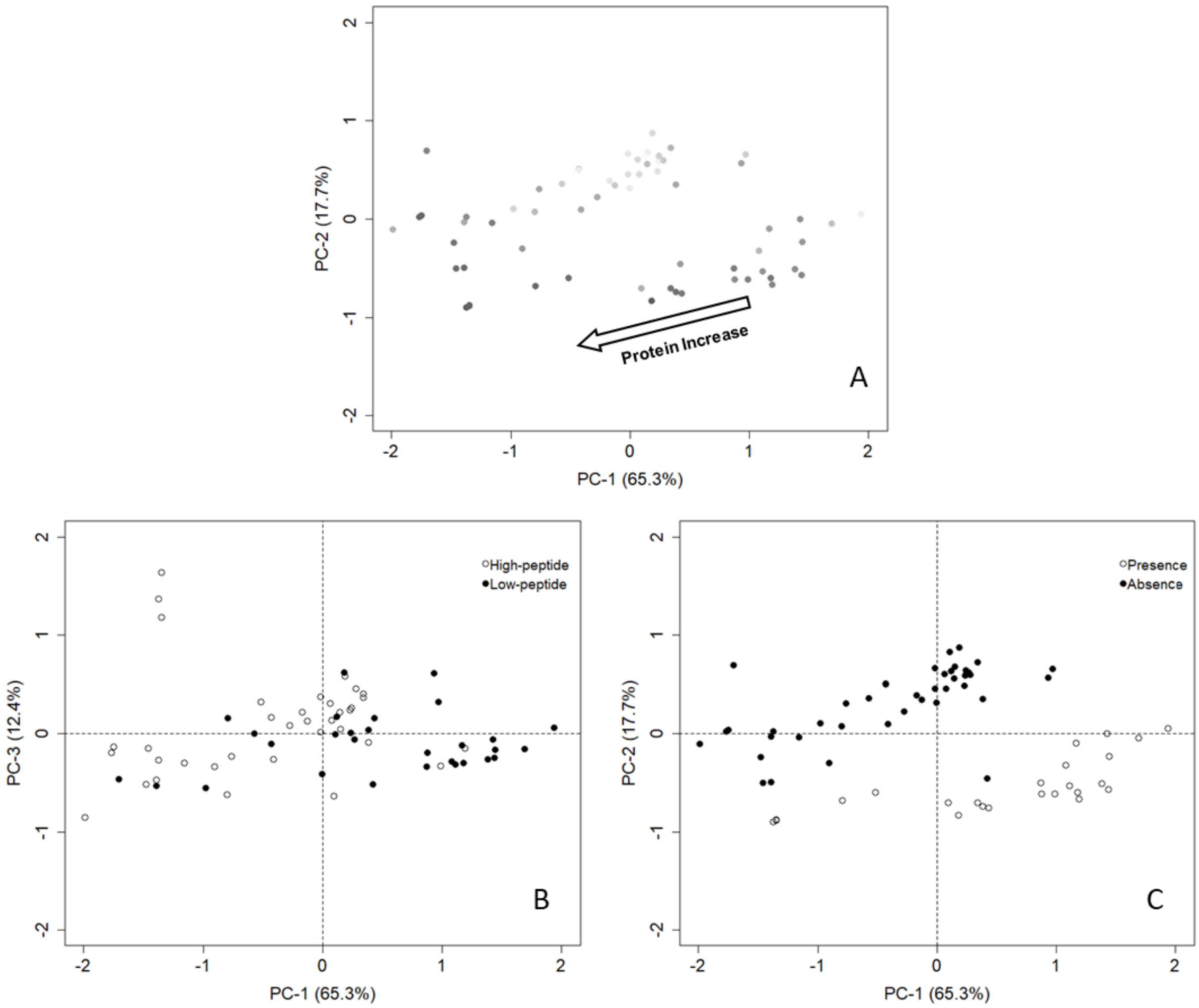


Figure 6-8 Principal component analysis similarity plot of A: high- and low-protein permeate the darker the shade indicates the increase of true protein content; B: high- and low-peptide permeate; C: presence or absence of full whey proteins (α-LA and β-LG) in permeate samples.

In Figure 6-8C, PC1 and PC2 provided visualization of the presence and absence of α-LA or β-LG among the permeate samples. Overall, PCA was able to distinguish the permeates with full protein from those that only contain fragments of protein or peptides. The fluorescence signal difference for full protein presence can be related to the quenching of tryptophan, and according to Chen and Barkley (1998), the side chain of the tryptophan can alter the fluorescence signal in proteins. In this study, fluorescence spectroscopy can distinguish the α-LA or β-LG corresponding side chains difference from the rest of the proteinaceous matters. The clear separation of the two groups also allows supervised classification for detection of α-LA or β-LG presence in permeate samples.

Partial least squares regression

PLSR was used to predict TP content in permeate, and Figure 6-9 provides the visualization for the prediction accuracy with external validation. The calibration model was established using 6 latent variables, and the RMSECV of the model was 0.20% (Figure 6-9A). An RPD of 2.6 was found from the calibration model, and according to Williams and Norris (2001), an RPD above 2 provides some practical applications for quality control purposes. The external validation yielded a RMSEP of 0.22 and an RPD of 2.8, showing that the calibration model is robust to test new samples (Figure 6-9B). The model developed using tryptophan excitation spectra can provide rapid estimation to the UF membrane processing industry. Kulmyrzaev et al. (2005) quantified the specific fraction of milk protein using fluorescence and PLSR, but the predictions were only useful for alkaline phosphatase and β-LG. In the present study, true protein value measured by Kjeldahl method seeks for less specificity but more accuracy towards estimating the overall proteinaceous matter in the permeate samples.

Partial least squares discriminative analysis

Since quantification may not be necessary when monitoring protein leak during UF, classification methods provides simpler outcomes and easy-to-interpret results. In Table 6-1, the LTP and HTP permeate are predicted and cross validated using PLS-DA. The accuracy of the model was 83.3% with sensitivity of 0.75 and specificity of 0.90. The sensitivity showed the model’s ability to successfully distinguish true HTP among all samples, and the specificity indicates the probability to confirm true LTP samples. Since the specificity is greater than the sensitivity, the classification error leads towards to predict false positives, meaning misclassified LTP to HTP samples (Ballabio and Consonni, 2013). Overall, the PLS-DA model provides useful prediction accuracy for quality control purposes. The flagged potential HTP products can be further tested with reference methods to confirm the protein leaks during UF.

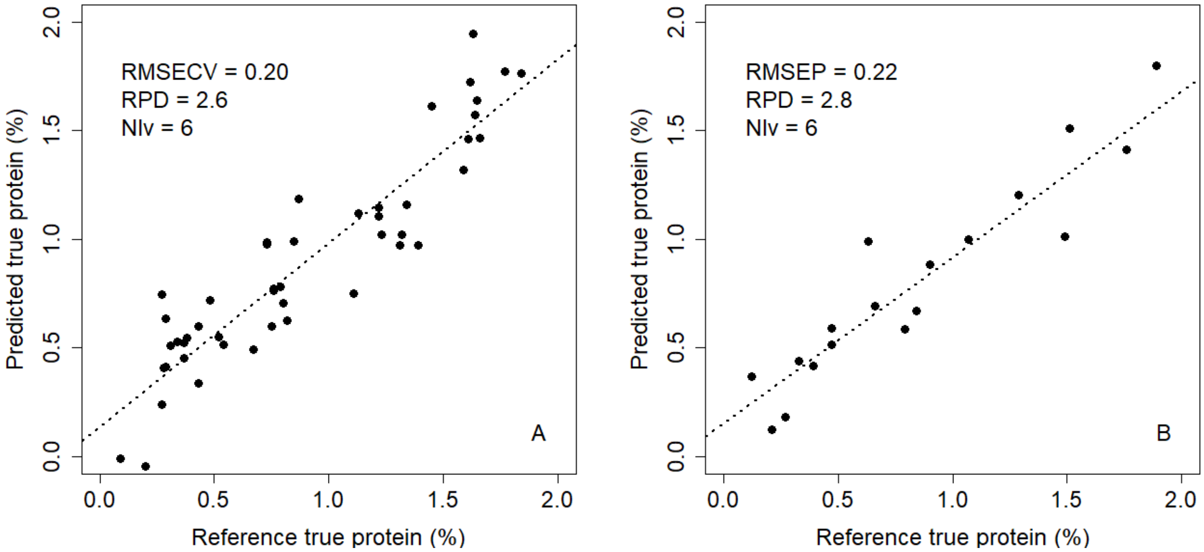


Figure 6-9 A: calibration model: reference true protein value vs. partial least square regression predictions of permeate samples. B: validation model: reference true protein value vs. partial least square regression predictions of permeate samples. The dashed line indicates the least-square fit of the predictions. Nlv = number of latent variables, RMSECV/P = root mean square error of cross validation/prediction, RPD = ratio of prediction to deviation.

To examine the specific leaked proteinaceous matters, Table 6-1 provides the predictions for tryptone-equivalent peptide contents and full whey protein presence. The accuracy for tryptone-equivalent peptide content classification (low vs. high) was 84.8% with a sensitivity of 0.91 and specificity of 0.78. The tryptone-equivalent peptide content prediction can serve as an indicator to the quality of the UF feed, and possible proteolysis from bacteria and enzyme residuals may lead to the high peptide content in permeates. Unlike the TP prediction, the specificity of the model is lower than the sensitivity, meaning that the model error comes from falsely classifying high-peptide samples as low-peptide, resulting in type II errors (Ballabio and

Consonni, 2013). Connecting to the tryptone-equivalent peptide discussion, the reference method for measuring peptide can be improved in the future to fully characterize the presence of high peptide content in permeates. The present method can successfully recognize permeates with α-LA or β-LG presence with an accuracy of 98.5%, sensitivity of 1.0, and specific of 0.98. The results agreed with previous studies of fluorescence spectroscopy characterizing proteins in dairy foods (Babu and Amamcharla, 2018). The model is capable of monitoring full protein presence which can be an indicator to the membrane processing quality. This accurate method to measure full protein presence can help the manufacturers monitor the UF processing and control production quality.

Table 6-1 Sensitivity, specificity, and classification error of partial least squares discriminant analysis models for true protein, peptide and full whey protein presence.

Classification	Class	Sensitivity	Specificity	Accuracy (%)
True Protein	Low/High	0.75	0.90	83.3%
Peptide	Low/High	0.91	0.78	84.8%
Full whey protein	Presence/Absence	1.00	0.98	98.5%

Conclusion

In this study, protein leak in UF permeates of whey and milk was characterized by various proteinaceous matters (e.g. TP, NPN, tryptone-equivalent peptide, α-LA, and β-LG). FFFS revealed available fluorophores in UF permeates and identified tryptophan as the fluorophore of interest for protein leak detection during UF. Chemometric analysis of the tryptophan excitation spectra classified permeates based on their TP, tryptone-equivalent peptide, and α-LA or β-LG presence. The cross-validated prediction methods yielded useful quantification of TP in permeates, and the classification methods was able to detect high TP level, high tryptone-equivalent peptide level, and presence of α-LA or β-LG with 83.3%, 84.8%, and 98.5% accuracy, respectively. The FFFS-based method can be further adopted to at-line and on-line modules in processing plants and provide real-time monitoring of protein leak in UF permeates.

References

ADPI (American Dairy Products Institute). 2018. Dairy permeate standard. Accessed Oct.15, 2018. https://www.adpi.org/Portals/0/Standards/DairyPermeateStandard_book.pdf

AOAC International. 2016. Official methods of analysis. 20th ed. AOAC International. Gaithersburg, MD

Andersen, C.M., and G. Mortensen. 2008. Fluorescence spectroscopy: a rapid tool for analyzing dairy products. J. Arg. Food Chem. 56:720–729.

Babu, K. S., and J. K. Amamcharla. 2018. Application of front-face fluorescence spectroscopy as a tool for monitoring changes in milk protein concentrate powders during storage. J. Dairy Sci.

Baker, A. 2005. Thermal fluorescence quenching properties of dissolved organic matter. Water Res. 39:4405–4412.

Ballabio, D., and V. Consonni. 2013. Classification tools in chemistry. Part 1: linear models. PLS-DA. Anal. Methods 5:3790–3798.

Birlouez-Aragon, I., P. Sabat, and N. Gouti. 2002. A new method for discriminating milk heat treatment. Int. Dairy J. 12:59–67.

Blecker, C., J.-M. Habib-Jiwan, and R. Karoui. 2012. Effect of heat treatment of rennet skim milk induced coagulation on the rheological properties and molecular structure determined by synchronous fluorescence spectroscopy and turbiscan. Food Chem. 135:1809–1817.

Bonfatti, V., L. Grigoletto, A. Cecchinato, L. Gallo, and P. Carnier. 2008. Validation of a new reversed-phase high-performance liquid chromatography method for separation and quantification of bovine milk protein genetic variants. J. Chromatogr. A 1195:101–106.

Chen, Y., and M.D. Barkley. 1998. Toward understanding tryptophan fluorescence in proteins. Biochemistry 37:9976–9982.

Diez, R., M.C. Ortiz, L. Sarabia, and I. Birlouez-Aragon. 2008. Potential of front face fluorescence associated to PLS regression to predict nutritional parameters in heat treated infant formula models. Anal. Chim. Acta 606:151–158.

Fagan, C.C., T.G. Ferreira, F.A. Payne, C.P. O'Donnell, D.J. O'Callaghan, and M. Castillo. 2011. Preliminary evaluation of endogenous milk fluorophores as tracer molecules for curd syneresis. J. Dairy Sci. 94:5350–5358.

Fox, P.F. 1989. Proteolysis during cheese manufacture and ripening1. J. Dairy Sci. 72:1379–1400.

Frankowski, K.M., R.E. Miracle, and M.A. Drake. 2014. The role of sodium in the salty taste of permeate. J. Dairy Sci. 97:5356–5370.

Karoui, R., and C. Blecker. 2011. Fluorescence spectroscopy measurement for quality assessment of food systems—a review. Food Bioprocess Tech. 4:364–386.

Karoui, R., G. Downey, and C. Blecker. 2010. Mid-infrared spectroscopy coupled with chemometrics: a tool for the analysis of intact food systems and the exploration of their molecular structure−quality relationships − a review. Chem. Rev. 110:6144–6168.

Karoui, R., A.M. Mouazen, E. Dufour, R. Schoonheydt, and J. De Baerdemaeker. 2006. Utilisation of front-face fluorescence spectroscopy for the determination of some selected chemical parameters in soft cheeses. Le Lait 86:155–169.

Kuhn, M. 2008. Caret package. J. Stat. Softw. 28: 1-26.

Kulmyrzaev, A.A., D. Levieux, and É. Dufour. 2005. Front-face fluorescence spectroscopy allows the characterization of mild heat treatments applied to milk. Relations with the denaturation of milk proteins. J. Agr. Food Chem. 53:502–507.

Lenhardt, L., R. Bro, I. Zeković, T. Dramićanin, and M.D. Dramićanin. 2015. Fluorescence spectroscopy coupled with PARAFAC and PLS DA for characterization and classification of honey. Food Chem. 175:284–291.

Li, B., H. Wang, Q. Zhao, J. Ouyang, and Y. Wu. 2015. Rapid detection of authenticity and adulteration of walnut oil by FTIR and fluorescence spectroscopy: A comparative study. Food Chem. 181:25–30.

Liu, X., and L.E. Metzger. 2007. Application of fluorescence spectroscopy for monitoring changes in nonfat dry milk during storage. J. Dairy Sci. 90:24–37.

Mevik, B.H., and R. Wehrens. 2007. The pls package: principal component and partial least squares regression in r. J. Stat. Softw. 018.

Minervini, F., F. Algaron, C.G. Rizzello, P.F. Fox, V. Monnet, and M. Gobbetti. 2003. Angiotensin i-converting-enzyme-inhibitory and antibacterial peptides from lactobacillus helveticus pr4 proteinase-hydrolyzed caseins of milk from six species. Appl. Environ. Microbiol. 69:5297–5305.

Oh, N.S., H.A. Lee, J.Y. Lee, J.Y. Joung, K.B. Lee, Y. Kim, K.W. Lee, and S.H. Kim. 2013. The dual effects of Maillard reaction and enzymatic hydrolysis on the antioxidant activity of milk proteins. J. Dairy Sci. 96:4899–4911.

R Core Team. 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Royer, C.A. 2006. Probing Protein Folding and Conformational Transitions with Fluorescence. Chem. Rev. 106:1769–1784.

Shaikh, S., and C. O'Donnell. 2017. Applications of fluorescence spectroscopy in dairy processing: a review. Curr. Opin. Food Sci. 17:16–24.

Sithole, R., M.R. McDaniel, and L.M. Goddik. 2005. Rate of Maillard browning in sweet whey powder. J. Dairy Sci. 88:1636–1645.

Smith, S.T., L. Metzger, and M.A. Drake. 2016. Evaluation of whey, milk, and delactosed permeates as salt substitutes. J. Dairy Sci. 99:8687–8698.

Suárez, E., A. Lobo, S. Alvarez, F.A. Riera, and R. Álvarez. 2009. Demineralization of whey and milk ultrafiltration permeate by means of nanofiltration. Desalination 241:272–280.

Vivian, J.T., and P.R. Callis. 2001. Mechanisms of tryptophan fluorescence shifts in proteins. Biophysic. J. 80:2093–2109.

Udenfriend, S., S. Stein, P. Böhlen, W. Dairman, W. Leimgruber, and M. Weigele. 1972. Fluorescamine: a reagent for assay of amino acids, peptides, proteins, and primary amines in the picomole range. Science 178:871–872.

Walstra, P., T. J. Geurts, A. Noomen, A. Jellema, and M. A. J. S. Van Boekel. 2005. Chapter 1: Milk – composition, structure, and properties. Pages 1–24 in Dairy technology: principles of milk properties and processes. Marcel Dekker Inc., New York, NY.

Williams, P., and K. Norris. 2001. Near-infrared technology in the agricultural and food industries, 2nd ed. P. Williams and K. Norris, ed. AACC, St. Paul, MN.

Wold, S., M. Sjöström, and L. Eriksson. 2001. PLS-regression: a basic tool of chemometrics. Chemometrics and Intelligent Laboratory Systems 58:109–130.

Chapter 7- Conclusions

Front-face fluorescence spectroscopy (FFFS) and chemometrics offer rapid and sensitive analysis in dairy foods. Using intrinsic fluorophores such as tryptophan, chemical properties of dairy foods can be characterized based on the FFFS spectra. Chemometric modelling of the obtained spectra extracts useful information for quantification and classification of dairy foods in a predictive manner. Chapters 4 and 5 developed a FFFS-based casein quantification method that can serve the cheese and fluid milk industry. Based on external validations, this method showed that:

- FFFS combined with chemometrics provides useful quantification of casein in raw, skim, and ultrafiltrated milk with a root mean square error ranged from 0.12 to 0.15% and relative error of 4.9 to 6.7% compared to the reference method.
- FFFS combined with partial least squares regression (PLSR) can estimate the casein-to-crude-protein ratio with a relative error of 3.0%.
- The practical prediction power from the FFFS-based method suggested that it can be implement as a rapid measurement for routine casein quantification in the dairy industry.

Chapter 6 applied FFFS and chemometrics to characterize protein leak during of skim milk and whey. Using both non-target and target FFFS approaches, this study concludes that:

- Full whey proteins (α-LA and β-LG) and tryptone-equivalent peptides were present in commercial UF permeates.
- Tryptophan was selected as the fluorophore of interest for protein leak detection in UF permeate.

- PLSR of the tryptophan excitation spectra quantified true protein content in liquid permeate stream with practical model performance for quality control purposes.
- Partial least squares discriminant analysis (PLSDA) of the tryptophan excitation spectra identified the source of protein leak and detected high true protein content, high tryptone-equivalent peptide content, and presence of α-LA or β-LG with 83.3%, 84.8%, and 98.5% cross-validated accuracies.

Appendix A - R code for chapter 4

```
#packages needed#
install.packages("caret")
install.packages("pls")
install.packages("clusterSim")
install.packages("prospectr")


library(caret)
library(pls)
library(clusterSim)
library(prospectr)


#data loading and explorotory analysis#
rawmilk<-read.csv(file="....csv", header= T)
View(rawmilk)
dim(rawmilk)
#Emission maxima#
max<-vector(length = 20)
for (i in 1:20){
 max[i]<-max(rawmilk[i,4:281])
}
print(max)
 cols<-character(length = 20)
for (j in max){
  cols[1:20]<-which(rawmilk == j, arr.ind = T)
}
cols
max[1]
which(rawmilk == 235.69, arr.ind = T)

#Spectra normalization and find derivatives#
rawmilk.std<-data.Normalization(rawmilk[,-
c(1:5,203:282)],type="n0",normalization = "row")

# p = polynomial order w = window size (must be odd) m = m-th derivative
(0
# = smoothing) The function accepts vectors, data.frames or matrices. For
a
# matrix input, observations should be arranged row-wise
SG.1st<- savitzkyGolay(rawmilk.std, p = 2, w = 19 , m = 1)

#Principle component analysis and visualization#
pcrm<-prcomp(SG.1st)
pcscores<-pcrm$x
PCAlab<-as.factor(rawmilk$protein)
plot(pcrm$x[,1], pcrm$x[,2], pch = 19, lwd = 4,cex.lab =1.5,cex.axis =
1.3,
     xlab="PC1 - 99.0%",ylab="PC2 - 0.3%")
abline(h=0, v=0, lty = 2)
text(pcrm$x[,1], pcrm$x[,2], labels = PCAlab, cex=1.4, pos = 2)
```

```
#model building and calibration visualization#
rawmilk.normed<-cbind(rawmilk$Casein, SG.1st)
colnames(rawmilk.normed)[1]<-"protein"
rawmilk.normed<-as.data.frame(rawmilk.normed)
tc<-trainControl(method = "LOOCV") #tune parameters#
plsrm<-train(protein ~.,  data = rawmilk.normed, tuneLength = 10, method =
"pcr", trControl = tc)
plsrm
predics<-as.data.frame(plsrm$pred)
plot(ncomp1$pred,ncomp1$obs, main = "GS.1st with PLS 22 variables")
abline(lm(ncomp1$pred~ncomp1$obs))
text(ncomp1$pred, ncomp1$obs, labels = rawmilk$Sample)
#RPD#
sd(rawmilk.normed$protein)/0.697

#validation#
rawval<-read.csv(file="....csv", header= T)

rawval.std<-data.Normalization(rawval[,-c(1,3)],type="n0",normalization =
"row")
rawval.all<- savitzkyGolay(rawval.std, p = 2, w = 19, m = 1)
rawval.all<-as.data.frame(rawval.all)
dim(rawval.all)
preds<-predict(plsrm$finalModel, newdata = rawval.all)
preds
plot(preds$`.outcome.1 comps`,rawval$Protein)
```

Appendix B - R code for chapter 5

```
#############packages################
install.packages("caret")
install.packages("pls")
install.packages("clusterSim")
install.packages("prospectr")
install.packages("glmnet")

library(clusterSim)
library(prospectr)
library(pls)
library(caret)
library(glmnet)

############evaluation functions#########
rmse<-function(x,y){
  sqrt(mean((x-y)^2))
}
bias<-function(x,y){
  mean(abs(x-y))
}
sep<-function(x,y){
  (sqrt(mean((x-y)^2)))^2-(mean(abs(x-y)))^2
}

rsq <- function(x, y){cor(x, y) ^ 2}

rpd<-function(x, y){
  sd(x)/rmse(x,y)
}

rpe <- function (x,y){
  rmse(x,y)/mean(x)*100
}

diag<-function(x,y){
  result<-cbind(rmse(x,y), bias(x,y), rsq(x,y), rpd(x,y), rpe(x,y))
  colnames(result)<-c("RMSE", "bias","R-2", "RPD", "RPE")
  options(digits = 4)
  return(result)
}
##########load data#################
cal<-read.csv(file = "...csv", header = T)
dim(cal)
val<-read.csv(file = "...csv", header = T)
dim(val)

#########preprocessing#######
calsg<-savitzkyGolay(cal[,-c(1:4)], p = 2, w = 19 , m = 1)
calnorm<-data.Normalization(calsg, type = "n4", normalization = "row")
valsg<-savitzkyGolay(val[,-c(1:7)], p = 2, w = 19 , m = 1)
valnorm<-data.Normalization(valsg, type= "n4" ,normalization = "row")
```

```
#########data partition###############
set.seed(2007)
index<-createDataPartition(val$CN,times = 1, p = 0.5, list = F)
val.train<-val$CN[index]
val.test<-val$CN[-index]
valsg.train<-as.data.frame(valsg[index,])
valsg.test<-as.data.frame(valsg[-index,])

#########PLSR calibration model training#########
pls<-plsr(cal$ratio~calsg, ncomp = 10, validation = "none")
summary(pls)
pre<-as.data.frame(pls$fitted.values)

#########validations and diagnostics ###############
error<-data.frame(matrix(nrow = 10, ncol = 5))
for (i in 1:10){
  preds<-predict(pls, ncomp = i, as.matrix(valsg.train))
  error[i,]<-diag(val.train, preds)
}
error

preds<-predict(pls, ncomp = which.min(error$X1) , as.matrix(valsg.train))
diag(val.train, preds)

lcor<-lm(val.train~preds)
lcor

#########test the final model###############
predscor<-predict(pls, ncomp = which.min(error$X1), as.matrix(valsg.test))
new<-predscor*0.641+0.270 #taken from the slope and intercept from lcor#
diag(val.test,new)


#########Elastic Net Regression#########
set.seed(2018)

diagnosis<-data.frame(matrix(nrow = 10, ncol = 5))
for (i in 0:10){
  fit <- cv.glmnet(calsg, cal$ratio, type.measure="mse",
                   alpha=(i/10), family="gaussian")
  predsen<-predict(fit, s = fit$lambda.1se,
                   newx = as.matrix(valsg.train))
  diagnosis[i,]<-diag(val.train, predsen)
}

diagnosis

set.seed(2018)
fit <- cv.glmnet(calsg, cal$ratio, type.measure="mse",
                 alpha=((which.min$X1)/10), family="gaussian")
val.en<-predict(fit, s = fit$lambda.1se,
                         newx = as.matrix(calsg))
diag(cal$ratio, val.en)
```

```
predsen<-predict(fit, s = fit$lambda.1se,
                 newx = as.matrix(valsg.train))
predsen

lcor.en<-lm(val.train~predsen)
lcor.en

predsen.test<-predict(fit, s = fit$lambda.1se,
                 newx = as.matrix(valsg.test))

new.en<-predsen.test*0.530+0.326 #taken output from lcor.en#
diag(val.test, new.en)
```

Appendix C – R code for chapter 6

```
###########Loading library################
install.packages("caret")
install.packages("pls")
install.packages("clusterSim")
install.packages("prospectr")

library(clusterSim)
library(prospectr)
library(pls)
library(caret)

############evaluation functions#########
rmse<-function(x,y){
  sqrt(mean((x-y)^2))
}
bias<-function(x,y){
  mean(abs(x-y))
}
sep<-function(x,y){
  (sqrt(mean((x-y)^2)))^2-(mean(abs(x-y)))^2
}

rsq <- function(x, y){cor(x, y) ^ 2}

rpd<-function(x, y){
  sd(x)/rmse(x,y)
}

rpe <- function (x,y){
  rmse(x,y)/mean(x)*100
}

diag<-function(x,y){
  result<-cbind(rmse(x,y), bias(x,y), rsq(x,y), rpd(x,y), rpe(x,y))
  colnames(result)<-c("RMSE", "bias","R-2", "RPD", "RPE")
  options(digits = 4)
  return(result)
}


#######PCA and normalization#######
Ex<-read.csv(file="C:/GradSchool/Grad research/Membrane processing
Fluorescence Study/excitation spectra/Excitation.csv", header = T)
Exmean<-data.Normalization(Ex[,14:209],type="n9",normalization = "row")
pc<-prcomp(Exmean, center = T, scale = T)
summary(pc)

#######plsr protein, 70/30 Split######
set.seed(2003)
index<-createDataPartition(Ex$Protein,times = 1, p = 0.7, list = F)
Ex.train<-Exmean[index,]
Ex.test<-Exmean[-index,]
EP.train<-Ex$Protein[index]
EP.test<-Ex$Protein[-index]

pls<-plsr(EP.train~ as.matrix(Ex.train), ncomp = 10, validation = "LOO")
summary(pls) #find the optimal number of components using the lowest
RMSECV#
diag(EP.train, cal$`EP.train.6 comps`)
pre<-predict(pls,ncomp = 6, newdata = as.matrix(Ex.test))
diag(EP.test, pre)

##########classification protein##############
proclass<-cbind(Ex$Proteincat,Exmean)
colnames(proclass)[1]<-"proteincat"
plsdapro<-train(proteincat ~., data = proclass,tuneLength = 10, method =
"pls", trControl = tc)
plsdapro
daresult<-as.data.frame(plsdapro$pred)
ncomp <- subset(daresult, ncomp == 5)
xtabs(~ncomp$obs+ncomp$pred)

##########plsda peptide##############

pepclass<-cbind(Ex$pepcat,Exmean)
colnames(pepclass)[1]<-"pepclass"
plsdapep<-train(pepclass ~. , data = pepclass, tuneLength = 10, method =
"pls", trControl = tc)
plsdapep
daresult<-as.data.frame(plsdapep$pred)
ncomp <- subset(daresult, ncomp == 5)
xtabs(~ncomp$pred+ncomp$obs)

##########plsda mppvsdpw############

sampleclass<-cbind(Ex$Spcat,Exmean)
colnames(sampleclass)[1]<-"samplecat"
plsdasp<-train(samplecat ~. , data = sampleclass, tuneLength = 10, method
= "pls", trControl = tc)
plsdasp
daresult<-as.data.frame(plsdasp$pred)
ncomp <- subset(daresult, ncomp == 2)
xtabs(~ncomp$pred+ncomp$obs)
```