MEASUREMENT OF CASEIN CONTENT IN MILK AND MILK-BASED PRODUCTS USING KJELDAHL, SODIUM DODECYL SULFATE POLYACRYLAMIDE GEL ELECTROPHORESIS, AND MID INFRARED SPECTROSCOPY

A Thesis

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Master of Science

by

Joice Pranata

December 2020

ABSTRACT

The objective of the first part of the study was to develop and validate partial least squares (PLS) models using MIR spectra to predict true protein (TP), casein (CN), serum protein (SP), and casein as a percentage of TP (CN%TP) content of micellar casein concentrates (MCC) and unflavored milk-based beverages. The PLS models, developed using 625 samples of formulated milk with a range of CN%TP and orthogonal matrix of fat, TP, and lactose were used for model development based on Kjeldahl reference chemistry and mid-infrared spectra collected for each sample. The ratio of performance to deviation (RPD) values were 62.7, 31.4, 14.1, and 3.8 for the prediction of TP, CN, SP, and CN%TP, respectively. External validation of the PLS models using 18 formulated MCC and 28 formulated unflavored milk-based beverage produced good validation performance, which may be attributed to the high RPD values achieved by orthogonal design of the modeling population used for the PLS model development.

The objectives of the second part of the study were to determine if milk CN%TP estimated by sodium dodecyl sulfate polyacrylamide gel (SDS-PAGE) is equivalent to CN%TP estimated by Kjeldahl, and to determine the proportion of CN, casein proteolysis products (CNPP), and SP from milk TP that goes into the Kjeldahl non-casein nitrogen (NCN) filtrate and the proportion that stays in the NCN precipitate using SDS-PAGE. The estimate of CN%TP by Kjeldahl was higher than the estimate of CN%TP by SDS-PAGE that was calculated as only intact CN divided by the total of all protein bands. However, no difference was detected in the estimate of CN%TP by Kjeldahl compared to CN%TP by SDS-PAGE when CNPP were included as CN in the calculation of SDS-PAGE results. SDS-PAGE analysis of the Kjeldahl NCN filtrate and NCN precipitate of three formulated milks with varying fat, lactose, TP, CN, and SP

content showed that a majority (89%) of the CNPP from the milk (approximately 10.13% out of 11.41% TP) were retained in the Kjeldahl NCN precipitate. In conclusion, the CN%TP measured by Kjeldahl underestimates the amount of proteolytic damage that has been done to casein in milk.

BIOGRAPHICAL SKETCH

Joice Pranata was born on June 2nd, 1993 in Jakarta, Indonesia. Joice attended Cornell University for her undergraduate studies where she majored in Food Science and minored in Applied Economics. During her undergraduate studies, she assisted in research on effects of high pressure processing on milk proteins in the Moraru lab, which specializes in dairy processing. She also participated in the IFTSA MARS product development competition and was a member of the Cornell team that earned honorable mention in the 2015 IFT annual meeting. Joice graduated in 2015 with a B.S. degree in Food Science and a summa cum laude distinction. She also earned the Professor C. Y. Lee Excellence in Food Chemistry Award and was recognized as a 2015 Merrill Presidential Scholar.

Following her graduation, Joice did a summer internship in operations research at Great Lakes Cheese in Cuba, NY under the supervision of Dr. Barbano. She then joined Lux Research Inc. as a Research Associate for the Food and Nutrition division where she assisted industry clients in conducting market analysis and technology scouting.

In 2018, she returned to Cornell University to pursue an M.S. in Food Science with a concentration in Food Chemistry. As a graduate student, Joice continued her participation as a team member in the IFTSA MARS Product Development Competition, and her team placed third in 2019 and first in 2020. She also participated as team captain in the National Dairy Council Product Competition and her team placed second in the 2019 ADSA annual meeting. Upon graduation, Joice will continue to pursue completion of a PhD degree under the Barbano Lab.

To my mother, Linda Widjaja.

ACKNOWLEDGMENTS

My deepest gratitude goes to my advisor, Dr. David Barbano, who has always patiently guided me in my academic pursuits and graciously poured his time and energy in mentoring me to become a better researcher. His passion and enthusiasm for research have been the source of inspiration that brought me to where I am today.

I sincerely thank Dr. Jura Liaukonyte for her constant support in her role as a committee member and minor advisor.

I am grateful to the Northeast Dairy Foods Research Center and the Cornell University Department of Food Science for their funding in this research.

I would like to express my immense appreciation for lab members: Chassidy Coon, Michelle Billota, Sara Hatch, and Ashley Kosicky for their technical assistance.

I would also like to specifically thank Larissa DiMarzo for her guidance as a fellow researcher and her primary role in the accomplishment of the research covered in this thesis.

Lastly, I would like to extend special thanks to my family for their constant love and support through every stage of my life as well as to the healthcare workers at Cornell for their dedication in providing care to our community.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

α_s-CN	Alpha-s-casein
α-LA	Alpha lactalbumin
β-CN	Beta-casein
β-LG	Beta lactoglobulin
ɣ-CN	Gamma-casein
κ-CN	Kappa-casein
AOACI	Association of Official Analytical Collaboration International
BSA	Bovine serum albumin
CF	Concentration factor
CN	Casein
CN%TP	Casein as a percentage of true protein
CNPP	Casein proteolysis products
d	readabilities
DTT	Dithiothreitol
FA	Fatty acid
FPRESS	F-Test predicted residual sum of squares
FTIR	Fourier Transform Infrared
IDF	International Dairy Federation
IGG	Immunoglobulin G
LFR	Lactoferrin
MCC	Micellar casein concentrates
MD	Mean difference
MF	Microfiltration
MIR	Mid infrared
Mr	Molecular mass

NCN	Non-casein nitrogen
NPN	Non-protein nitrogen
NPN%TP	Non-protein nitrogen content as a percentage of true protein
OD	Optical density
PCA	Principal component analysis
PLS	Partial least squares
PMO	Pasteurized Milk Ordinance
PPII	poly-L-proline II-like structures
PRESS	F-ratio predicted residual sum of squares
RO	Reverse osmosis
RPD	Ratio of performance to deviation
RSD	Relative standard deviation
SCC	Somatic cell count
SD	Standard deviation
SDD	Standard deviation of the difference
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SECV	Standard error of cross-validation
SEP	Standard error of prediction
SP	Serum protein
SPI	Serum protein isolate
TCA	Tricholoroacetic acid
TN	Total nitrogen
TP	True protein
UF	Ultrafiltration

CHAPTER ONE

INTRODUCTION

Milk Protein Analysis by Kjeldahl

Principle of the Kjeldahl Method. The Kjeldahl method is the reference method for milk protein determination (AOACI, 2019; method 991.20). This method determines the total nitrogen (**TN**) in milk, and the percent nitrogen in milk multiplied by a factor of 6.38 gives the percent total protein in milk (Barbano et al., 1990). The Kjeldahl method consists of three main steps: digestion, distillation, and titration (AOACI, 2019). In the digestion step, milk is digested under heat with sulfuric acid to release nitrogen as ammonium salt. In the distillation step, ammonium is reacted with sodium hydroxide to produce ammonia that is distilled and then collected in boric acid. In the titration step, the collected ammonium borate is neutralized with hydrochloric acid in the presence of an indicator, and total nitrogen is determined based on the volume of acid used in the titration.

Rowland Nitrogen Fractions in Milk. Rowland (1938) divided nitrogen in milk into several fractions, including total nitrogen, non-protein nitrogen (**NPN**), and non-casein nitrogen (**NCN**), and developed modified versions of the Kjeldahl method for the determination of these nitrogen fractions. Historically, milk protein content has been determined from TN, but this results in poor accuracy because approximately 5 to 6% of milk TN is not associated with protein and that NPN value varies across different milks (Barbano and Lynch, 1992). The NPN is measured from the nitrogen content of the soluble fraction of milk treated with 12% trichloroacetic acid (**TCA**)

(AOACI, 2019; method 991.21). The NCN is measured from the nitrogen content of the soluble fraction of milk acidified to pH 4.6 using acetic acid and sodium actetate solutions (AOACI, 2019; method 998.05). The TN subtracted by the NPN and multiplied by 6.38 gives the true protein (**TP**) content of milk. The TN subtracted by the NCN and multiplied by 6.38 is defined as the casein (**CN**) content of milk.

The Kjeldahl NPN method can be used to measure the TP content of milk based on the difference between the TN of the milk and NPN of the filtrate of the milk in 12% TCA solution (AOACI, 2019; method 991.21). Another method to determine TP that does not require two separate measurements for TN and NPN content is the direct TP method, which directly measures the nitrogen content of the precipitate of milk treated with 12% TCA (AOACI 2019, method 991.22). A collaborative study by Lynch et al. (1991) demonstrated that the direct TP method gave acceptable performance, and the method has been adopted by Association of Official Analytical Chemists International (**AOACI**).

The Kjeldahl NCN method indirectly measures the CN content of milk based on the difference between TN of the milk and NCN content of the filtrate of the milk acidified to pH 4.6 (AOACI, 2019; method 991.21). In the direct method for determination of CN, the nitrogen content of the precipitate of milk at pH 4.6 is measured by Kjeldahl analysis, thus requiring only one Kjeldahl analysis for CN instead of two Kjeldahl analysis for TN and NCN (AOACI, 2019; method 998.06). The direct CN method that was modified to have milk samples precipitated directly in the Kjeldahl flask showed acceptable performance and was adopted by AOACI (Lynch et al., 1998).

Casein as a Percentage of True Protein. The CN and TP values obtained from Kjeldahl analysis are used to calculate the value of casein as a percentage of true protein (**CN%TP**). On average, high quality raw bovine milk has a CN%TP value of 82%, and the CN%TP value fluctuates the most due to seasonal changes (Kindstedt et al., 1983). CN%TP value serves as an indicator for milk protein quality because its value decreases with increasing breakdown of CN into casein proteolysis products (**CNPP**) by proteolytic enzymes.

Verdi et al. (1987) found that CN%TP was significantly higher in milk with lower somatic cell count (**SCC**) for most of the studied cows' lactation period. SCC is one of the main indicators of milk quality, and a high SCC in milk indicates a cow may be suffering from mastitis (Andrews, 1983). Milks with high SCC tend to have higher degrees of proteolysis (Ma et al., 2000). Higher degrees of proteolysis usually translate to lower CN%TP in milk. Proteolysis in milk has been shown to result in astringent flavor (Harwalkar et al., 1993). In cheese, proteolytic damage to caseins caused by native milk proteases leads to lower curd strength, lower syneresis rate, and higher moisture content (Kelly et al., 1982; O'Keeffe, 1984). Degree of proteolysis can provide a reflection of the milk's flavor quality and potential functionalities when converted into other dairy products such as yogurt, cheese, and milk powder. Therefore, dairy farmers are incentivized to produce milk with low SCC values and low levels of casein proteolysis, which can be reflected by high CN%TP values. In the United States, the regulatory limit for SCC as defined by the Grade "A" Pasteurized Milk Ordinance (**PMO**) is 750,000 cells per ml.

Aside from providing a measure of proteolytic breakdown, CN%TP value, when modified by filtration technology to levels outside the biological variation range, also has influences on visual and textural sensory properties of milk-based beverages. A study by Cheng et al. (2019) showed that increasing CN%TP leads to increase in perceived whiteness and opacity and decrease in perceived color intensity and yellowness in unpasteurized milk-based beverages. A study by Misawa et al. (2016) also found that increasing CN%TP in milk leads to increased viscosity as well as increased whiteness and decreased yellowness as measured by higher L and lower b value, respectively, on the Hunter L, a, b color scale.

Casein Proteolysis Products in Kjeldahl NCN Method. Based on the NCN Kjeldahl method, casein content in milk is defined as all milk proteins that are precipitated at pH 4.6. However, this may include serum proteins (**SP**) and proteolysis products of CN that precipitate with intact CN proteins. CN content can also be more accurately defined as only intact CN proteins in milk, which includes α_s-casein (**α_s-CN**), β-casein (**β-CN**), and κ-casein (**κ-CN**).

Karman and Van Boekel (1986) showed that during Kjeldahl NCN analysis, a portion of CNPP may remain in the NCN precipitate while the rest remain soluble in the filtrate. If CN content is defined as intact CN, this would lead to an overestimation of intact CN by the Kjeldahl NCN method. Portions of intact CN may also remain in the filtrate during Kjeldahl NCN analysis and contribute to an underestimation of CN. Zhang and Metzger (2011) showed through sodium dodecyl sulfate polyacrylamide gel (**SDS-PAGE**) analysis that filtrates of milk at pH 4.6 still contain intact caseins, particularly α_s-casein.

Milk Protein Analysis by Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis

Principle of SDS-PAGE. SDS-PAGE is a method to separate proteins based on their molecular masses under an electric field. SDS-PAGE of milk enables the quantification of relative proportions of CN, CNPP, and SP (Verdi et al, 1987). In the sample preparation, reducing agent dithiothreitol (**DTT**) breaks apart the protein's disulfide bonds, and sodium dodecyl sulfate (**SDS**), an anionic detergent, forms complexes with the polypeptide chains to produce a linear polypeptide chain, effectively denaturing the structure of the protein. SDS evenly imparts a negative charge throughout the polypeptide chain that is proportional to the molecular mass (**Mr**) of the polypeptide at 1.4 grams of SDS bound per gram of protein (Wrolstad et al., 2005). The SDS-protein complex has very high negative charge for all proteins, and the relative difference in total charge among different protein-SDS complexes is small. Thus, proteins separate in SDS-PAGE based on their difference in molecular weight and shape, but not charge. These pre-treatments on the sample remove most of the effect of the native proteins' structure and charge.

In the electrophoresis step, the treated protein samples are loaded onto a polyacrylamide gel and an electrical voltage is applied to the gel, causing the denatured proteins to migrate through the gel towards the positive electrode (anode). Polyacrylamide gels consist of long chains of polymerized acrylamide monomers that are cross-linked by N,N'-methylene bisacrylamide (bisacrylamide). These polymerized acrylamide chains form a mesh with pore sizes that depend on the total

acrylamide (acrylamide monomers and bisacrylamide) concentration and the proportion of bisacrylamide in the mixture. Increasing these results in gel matrices with smaller pore sizes, which limits the size of proteins that can pass through it. The speed at which proteins migrate through the gel matrix depends on the size of the molecule. Smaller (lower molecular weight) molecules migrate faster than larger molecules because they experience less resistance from the gel matrix, so the individual proteins can be separated relative to their size or molecular mass. Following electrophoresis, the gels are fixed and then stained with Coomasie Brilliant Blue dye to show the separated protein bands. The gels can then be scanned using a densitometer, and the relative densities of the dye bound to the protein bands can be used to calculate the relative concentration of the separated proteins (Wrolstad et al., 2005).

Quantification of Individual Milk Proteins by SDS-PAGE. Table 1.1 shows the molecular masses of the major milk proteins that appear on an SDS-PAGE gel. Figure 1.1 shows the relative positions of milk proteins on an SDS-PAGE discontinuous gel with 10-20% gradient. As can be seen from Figure 1.1 and Table 1.1, the milk proteins' relative position on the gel mostly follows the relative size of the proteins' molecular mass, with α-lactalbumin migrating the fastest. An exception to the relationship between protein mobility and protein size is with casein proteins. Generally, the SDS-protein complexes are globular in shape. However, milk caseins don't migrate strictly by molecular weight because of their high content of the amino acid proline in their primary amino acid sequence. The high frequency of proline in the amino acid sequence of casein monomers makes them resist being shaped into a

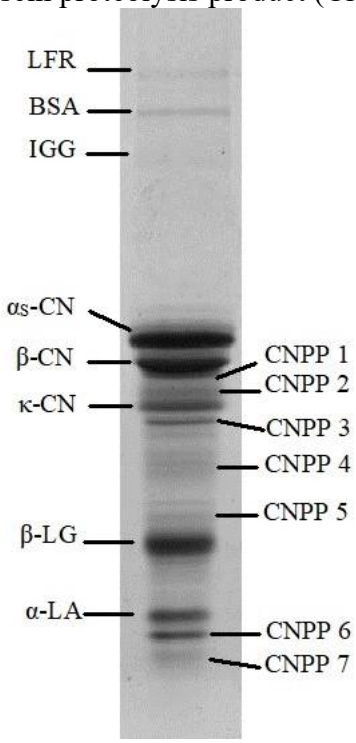
globlular (spherical) structure with the SDS. CN proteins, especially α_{s1}-casein, migrate at a slower rate than expected based on its molecular mass, and at pH 6.8, the pH of the sample buffer, the α_{s1}-CN and SDS complex has a larger size than the β-CN and SDS complex (Creamer and Richardson, 1984). Because a segment of the α_{s1}-CN polypeptide chain contain a high density of negatively charged amino acid side chains and high proline content, a large amount of electrostatic repulsion and steric hindrance occur, leading to α_{s1}-CN having a larger hydrodynamic size and slower electrophoretic mobility than expected from its molecular mass on an SDS-PAGE gel (Creamer and Richardson, 1984).

Table 1.1. Molecular masses of major bovine milk proteins that appear on SDS-PAGE electrophoresis.

Milk Protein	Molecular Mass (kDa)
lactoferrin	76.0[a]
bovine serum albumin	66.3[b]
immumoglobulin G (heavy chain)	56.0[c]
α_{s1}-casein	23.6[b]
α_{s2}-casein	25.2[b]
β-casein	24.0[b]
β-lactoglobulin	18.3[b]
α-lactalbumin	14.2[b]

[a]Castellino et al. (1970)
[b]Eigel et al. (1984)
[c]Bennett et al. (1994)

Figure 1.1. SDS-PAGE gel of milk showing the major milk proteins bands of bovine serum albumin (BSA), immunoglobulin G (IGG), lactoferrin (LFR), α_s-casein (α_s-CN), β-casein (β-CN), and κ-casein (κ-CN), β-lactoglobulin (β-LG), and α-lactalbumin (α-LA), as well as casein proteolysis product (CNPP) bands.



CNPP bands are also visible on SDS-PAGE gels, and their relative proportions are quantifiable. CNPP may result from CN proteolysis by native milk alkaline proteases, especially plasmin, or from bacterial proteases. In preserved milk, plasmin activity causes the majority of casein proteolysis, and the strongest absorption typically comes from CNPP 6, which appears after the α-lactalbumin band (Figure 1.1). Verdi et al. (1987) suggested that a large portion of the CNPP are ɣ-caseins (**ɣ-CN**), which are plasmin proteolysis products of β-CN, and that the CNPP 6 band consists of γ_2-CN and γ_3-CN. Amino acids 28 to 209 (Mr 20.5 kDa), 105 to 209 (Mr 11.8 kDa), and 107 to 209 (Mr 11.6 kDa) of β-CN form the CNPP γ_1-CN, γ_2-CN, and γ_3-CN, respectively (Eigel et al., 1984).

Milk Protein Analysis by Mid Infrared

Overview of Mid Infrared Milk Analyzers. In the dairy industry, infrared analyzers are used for the rapid analysis of milk components, especially for the quantification of fat, protein, lactose, and total solids. Mid infrared (**MIR**) milk analyzers are secondary testing instruments that are calibrated to agree with the results of reference chemistry methods.

The infrared spectral region ranges from 12800 to 10 cm^{-1} (0.78µm to 1000µm). This infrared region consists of the near infrared region from 12800 to 4000 cm^{-1} (780nm to 2.5µm), the MIR region from 4000 to 200 cm^{-1} (2.5µm to 50µm), and the far infrared region from 200 to 10 cm^{-1} (50µm to 1000µm). Chemical bonds of molecules rotate or vibrate at discrete energy levels, and these bonds transition from one state to another when they absorb energy that causes them to rotate or vibrate. Because molecular vibrations of chemical bonds absorb infrared light at specific frequencies, an infrared spectrum to evaluate the chemical composition of milk can be generated by measuring the milk's absorption of infrared light across a range of frequencies. The concentration of components in milk can be correlated to its infrared light absorbance, and this relationship is governed by the Beer-Lambert Law (Biggs et al., 1987).

The first generation of MIR milk analyzers measured the concentration of protein by the amide II stretch at 6.5 µm (1548 cm^{-1}), fat A by the carbonyl stretch at 5.7 µm (1745 cm^{-1}), fat B by the carbon-hydrogen stretch at 3.5 µm (2874 cm^{-1}), and lactose by carbon-hydroxyl stretch at 9.6 µm (1042 cm^{-1}) (Biggs et al., 1987). In early MIR instrumentation, a monochromator or an optical interference filter performs the

separation and selection of light wavelengths passing through. A monochromator disperses light consisting of a wide range of wavelengths into monochromatic lights at different wavelengths. MIR analyzers have used prisms and diffraction gratings as monochromators. Prisms in MIR analyzers are typically made of sodium chloride or potassium bromide. A dispersion grating consists of closely packed parallel grooves covered with reflective material. Light that hits the surface of the dispersion grating are dispersed into different wavelengths, each bouncing off at specific angles.

Optical interference filters consist of circular disks made of mineral crystals that select for specific range of wavelengths of MIR light. The range of wavelengths selected by the optical filter is referred to as band width and the mid-point of the band is called the center wavelength. Filter-based MIR analyzers usually contain four sample filters and four reference filters for wavelengths associated with protein, fat A, fat B, and lactose. In the first infrared milk analyzers, infrared light passed through a reference cell containing zeroing solution and sample cell containing milk sample with two parallel light paths and beam splitter that alternated light from the reference cell and the sample striking the detector (Goulden, 1961; Biggs, 1967). The reference signal serves to remove background noises from water absorption and light scattering. An infrared detector collects the absorption difference between sample and reference signals. Ultimately, as electronic circuit boards improved, the system was reduced to one cuvette and light path with the MIR light absorbance of the zeroing solution held in a memory circuit while the absorbance of the milk sample was measured. This ensured the cuvette path length was identical for the zeroing solution and the sample.

Modern MIR milk analyzers apply the principles of Fourier Transform Infrared (**FTIR**) spectroscopy. Instead of an optical interference filter or a monochromator, the FTIR instrument uses an interferometer and can make use of the entire MIR spectrum in the analysis of milk components. A Michelson interferometer consists of a fixed and a movable mirror. Light passes through a beam splitter and is split into two pathways with half reflecting onto the fixed mirror and the other half going to the movable mirror. The movable mirror shifts back and forth along an axis perpendicular to its plane to create variable optical path lengths for the light that is reflected by it. The two light rays with different path lengths that are reflected from the fixed and movable mirrors will recombine with constructive and destructive interferences and produce light of variable intensities depending on the optical path length differences. The variable intensities of lights absorbed by the detector are recorded by the instrument relative to the position of the movable mirror to produce an interferogram. The exact position is tracked and measured using a laser light source. The interferogram reflects the constructive and destructive interferences from the combination of the two light rays produced by the interferometer, and it is the Fourier transform of the spectrum.

A FTIR milk analyzer can be calibrated using a fixed-filter approach by creating virtual filters. For this approach, the center wavelengths and bandwidth for the virtual filters are set by the operator. Kaylegian et al. (2009) optimized virtual filters for a FTIR milk analyzer by minimizing intercorrection factors, and the optimized filter wavelength for measurement of the major components of milk are shown in Table 1.2. The Fat A sample virtual filter is the carbonyl stretch and the Fat

B sample virtual filter is the carbon hydrogen stretch in the structure of a triglyceride. The lactose sample virtual filter is the -OH stretch and protein sample virtual filter is the N-H stretch. A reference wavelength is paired with each of these sample wavelengths. The criteria for each reference virtual filter is that it is a band of wavelengths as close as possible to the sample filter and that when the concentration of the component of interest changes there is little if any change in light absorbance at the reference filter. Intercorrection factors are coefficients assigned to adjust for water displacement and absorbance by other components in milk at the target wavelength for the component being measured (Kaylegian et al., 2006) and typical intercorrection factors are also shown in Table 1.2. Thus, the intercorrection factors adjust spectral reading as a function of the concentration of the non-target component with which it is assigned. The advantage of virtual filters is that they can be set at exact wavelength and they are the same from one instrument to the next. This was not the case for the physical optical filters because their total transmission and bandwidth varied from filter.

Table 1.2. Optimized Fourier transform mid-infrared center wavelengths and bandwidths (i.e., minimized size of intercorrection factors). Reproduced from Kaylegian (2009).

	Wavelength (μm)				Intercorrection factor for secondary component			
	Sample		Reference					
Primary Component	Center	Bandwith	Center	Bandwith	Fat B	Lactose	Protein	Fat A
Fat B	3.508	0.032	3.556	0.030	1.000	-0.160	-0.065	–
Lactose	9.542	0.182	7.734	0.084	0.038	1.000	0.015	–
Protein	6.489	0.085	6.707	0.054	0.065	0.050	1.000	–
Fat A	5.721	0.052	5.583	0.050	–	0.030	0.025	1.000

Calibration of MIR milk analyzer for protein determination. MIR milk analyzers are secondary testing instruments for rapid determination of milk composition (AOACI, 2019; method 972.16). MIR milk analyzers need to be calibrated using milks that have known content determined using a reference chemistry method for each component being measured. For determination of protein, MIR analyzers are calibrated using Kjeldahl as reference chemistry method [AOACI, 2019; method 991.22 (TP) or 991.20 (TN) and 991.21 (NPN)].

As stated previously, the dairy industry has historically used the value of TN for determination of milk protein content. The use of TN instead of TP value for the calibration of MIR milk analyzers results in limited measurement accuracy as demonstrated by Barbano and Lynch (1992). They suggested that variations between laboratories in MIR readings for protein from the same milk sample are half due to variations in NPN content as a percentage of TN (**NPN%TP**) when TN value is used for calibration and half due to variation in Kjeldahl results obtained by the laboratories. They showed a systematic bias of up to 0.03 to 0.06% protein between laboratories and between months within the same laboratory caused by variations in NPN content as a percentage of TN. They also showed an even larger systematic bias of up to 0.1% protein due to variations in NPN%TP between milk samples within a set of calibration samples. The current use of TP value improves the accuracy of MIR protein determination by eliminating these biases.

Traditionally, MIR milk analyzers are calibrated using reference chemistry values from sample sets of preserved raw milk from individual producers. The actual

milk composition of samples from producer milks cannot be controlled and samples from producer milk will have a strong correlation between protein and fat concentrations, as well as a limited range in the concentration of components. Kaylegian et al. (2006) described a modified method for producing orthogonal calibration sample sets (modified milks) using combinations of pasteurized cream, ultrafiltration skim milk retentate, and permeate. They showed that using modified milks for the calibration of MIR milk analyzers can improve the instruments' performance because it can offer a larger range of concentration for each component, more uniform concentration distribution within the range, and lower correlation between fat and protein concentrations within a calibration set. Thus, the development of MIR models in this study uses modified milk samples.

Role of MIR instruments in the dairy industry. As an indirect testing method, MIR analyzers provide a rapid and cost-effective method for the measurement of milk components. In dairy processing, they are used for the determination of fat, protein, and lactose contents, which serve as parameters for milk payment to dairy farmers. In dairy farming, MIR instruments provide information on individual cow milk composition for efficient dairy herd management, such as in regulating diet forage concentration ratio, monitoring cow energy balance (McParland et al., 2014), and even detecting mastitis (Rienesl et al., 2019).

MIR instruments can also serve as useful tools for process control, such as for real-time checks in production lines to make sure dairy products are within specification, for standardization of milk to achieve a target fat and protein content after its separation into cream and skim milk, and for controlling the cheesemaking

process (Fagan et al., 2009). MIR instruments also have the potential for detecting adulteration in dairy products and checking product authenticity (Rodriguez-Saona and Allendorf, 2011).

Evolution in the application of MIR milk analyzers for measurement of fractions of milk components. MIR analyzers have potential applications for more detailed composition analysis of fractions of milk components, and this is especially true in the case of milk fat and its individual fatty acid (**FA**) composition. Currently, many models have been developed and evaluated for the prediction of milk FA content from MIR spectra. Partial least squares (**PLS**) regression is the common statistical method used for developing these models. The robustness of these models can be evaluated based on its ratio of performance to deviation (**RPD**) value. The RPD value is the standard deviation (**SD**) of reference values used in prediction or validation divided by the standard error of cross-validation (**SECV**). According to Manley (2014), a model with RPD higher than 3 can be used for screening, higher than 5 can be used for quality control, and higher than 8 can be used for any application.

Soyeurt et al. (2006) developed calibration equations for the prediction of C12:0, C14:0, C16:0, C16:1 cis-9, C18:1, saturated and monounsaturated fatty acids in milk. Upon evaluation using milk samples from different cow breeds, production systems, and countries, these calibration equations achieved an RPD of 3 or greater for all target components except for C16:1 cis-9 (Soyeurt et al., 2011). These calibration equations were updated using more samples in the European RobustMilk project, and the updated equations were evaluated by Maurice-Van Eijndhoven et al. (2012). These

calibration equations had an RPD value of greater than 3 for individual fatty acids C4:0, C6:0, C8:0, C10:0, C12:0, C14:0, C16:0, C17:0, and C18:0, and an RPD value of greater than 5 for saturated FA group, short chain FA group, and medium chain FA group. However, equations for individual FAs cis-C14:1 and cis-C16:1 had RPD values of less than 3.

Wojciechowski and Barbano (2016) later developed PLS models with RPD values of 2.1 and 3.3 for the determination of chain length and unsaturation of FAs, respectively. A PLS model for the determination of de novo (C4 to C4), mixed origin (C16:0, C16:1, C17:0), and preformed fatty acids (C$\geq$18:0) have also been developed and the models produced an RPD of greater than 5, with de novo FA group predictions achieving the highest RPD (Woolpert et al., 2016). Following the method for developing MIR models to quantify individual FA and FA groups in milk, it may be possible to also develop similar MIR models to quantify individual milk protein composition.

Principle of MIR Analysis of Protein. Proteins and peptides produce nine bands in the mid infrared range: Amide A, B, and I-VII (Carbonaro and Nucara, 2010). The amide I band around 1650 cm^{-1} arises predominantly from C=O stretch vibrations and to a lesser extent from N-H bend, C-N stretch, and CCN bend vibrations (Barth, 2007). The amide II band around 1550 cm^{-1} arises from the out of phase combination of the N-H bend and the C-N stretch vibrations (Barth, 2007). The amide III band around 1400 to 1200 cm^{-1} arises from the in-phase combination of the C-N stretch and the N-H stretch vibrations, and to a lesser extent from the C=O bend and C-C stretch vibrations (Barth, 2007). The amide A band of proteins at 3310 and

3270 cm^{-1} results from the NH stretch vibration, which also gives rise to a weak amide band between 3100 and 3030 cm^{-1} (Barth, 2007). These protein bands allow for the determination of protein concentration based on the absorption intensities of protein bands in MIR spectra. Milk proteins can be expected to exhibit unique molecular vibrations in the MIR range, and the differences in amino acid composition and protein structure of casein and serum proteins may provide the opportunity for predicting these individual proteins' compositions based on differences in their MIR absorbance characteristics.

Influence of Milk Protein Amino Acid Composition and Primary Structure on MIR spectra. Molecular bond vibrations in amino acids absorb infrared light of the same oscillation frequency. In infrared spectroscopy, the band position, band width, and absorption coefficient of the spectra are influenced by the structure of the amino acid side chains and the amino acids' environment (Barth, 2000). The protonation state, cation coordination, and hydrogen bonding of amino acid side chains strongly influence band position of the amino acid's absorbance (Barth, 2000). Increasing conformational flexibility of the molecule increases band width, and higher changes in dipole moment during vibration results in higher absorption coefficient (Barth, 2000). Environmental conditions such as the position of the amino acid in a protein and the pH of the solution also greatly influences the infrared absorption characteristics (Barth, 2000).

Vibrations from some amino acid side chains, especially arginine, glutamine, aspartic acid, glutamic acid, lysine, tyrosine, histidine, and phenyalanine may result in absorbance within the amide I region, although the contribution of these absorbances

are relatively low (Carbonaro and Nucara, 2010). Amino acid side chain vibrations have little influence on the amide I and II bands, but they have a strong influence on the amide III band (Barth, 2007).

Caseins, especially β-casein, are abundant in proline (Holt and Sawyer, 1993). Proline-rich proteins are expected to have a weakened amide II band intensity and a strong band arising from the C-N stretching vibration with a suggested band position around the 1450cm-1 region of the FTIR spectra that is largely influenced by water hydration state and hydrogen bonding (Banc et al., 2011). Compared to caseins, the whey proteins have higher levels of sulfur from cysteine. (Fox and McSweeney, 2003). In cysteine, the stretching of the S-H group is shifted to a band position that is free from overlap by other side chain vibrations due to the mass of the sulfur atom (Barth, 2000). β-lactoglobulin also contains a high amount of branched chain amino acids (Boland, 2011).

Influence of Milk Protein Secondary and Tertiary Structure on MIR spectra. The secondary structure and tertiary structure of proteins also shape the signals in the infrared spectra. The secondary structure of proteins has little influence on the amide A band (Barth, 2007). Secondary structures influence the amide I, II, and III bands of protein, but this is especially true for the amide I band, which is often used to predict the secondary structure composition of a protein (Barth, 2007). Within the amide I band, α-helical structures cause strong absorbance at around 1654 cm-1 (Carbonaro and Nucara, 2010). Antiparallel β-sheets give strong absorbance between 1612 and 1640 cm-1 and a weaker band around 1682 cm-1 and parallel β-sheets give absorbance between 1626 and1640 cm-1 (Carbonaro and Nucara, 2010). Random

coils tend to give absorbance around 1645 cm-1 (Carbonaro and Nucara, 2010). The α-helix band position shifts down with increasing helix length and short α-helices give rise to several bands throughout the amide I region (Carbonaro and Nucara, 2010). The β-sheets band position shifts down with increasing number of strands, and for antiparallel β-sheets, the position shifts up with twisting of the sheets (Carbonaro and Nucara, 2010).

In Fourier Transform Infrared (FTIR) spectra, the amide I band is centered around 1650 cm^{-1} for β-casein (Farrell et al., 2001) and around 1633 cm^{-1} for β-lactoglobulin (Bhattacharjee et al., 2005) when examined in H_2O under ambient conditions. Within the amide I region of the FTIR spectra of β-casein and α_{s1}-casein peptide (residues 1-23), the band at 1656-1658 cm-1 were assigned to large loops, which may also contain β-turns and α-helical structures (Farrell et al., 2001; Malin et al., 2001). In β-lactoglobulin, the band around 1658 cm-1 was attributed to α-helices (Bhattacharjee et al., 2005).

In αs1-casein N-terminal peptide and β-casein, peak areas at lower wavenumbers between 1620 to 1639 cm-1 were assigned to β sheets and short runs of poly-L-proline II-like structures (**PPII**), while peak areas at higher wavenumbers between 1661 to 1697 cm-1 were assigned to turns (Farrell et al., 2001; Malin et al., 2001). In β-lactoglobulin, the range of peak areas due to β -sheet was wider, covering between 1694 to 1626 cm-1, while the range of peak areas due to turns was narrower, covering between 1681 to 1661 cm-1 (Bhattacharjee et al., 2005).

Due to its high proline content, caseins tend to have less α-helix and β-sheet secondary structure compared to whey proteins, and conformation of caseins has been

described as rheomorphic (Holt and Sawyer, 1993). On the other hand, serum proteins tend to have more structured conformations. β-lactoglobulin has a globular structure with a goblet shape and a hydrophobic cavity (Boland, 2011) and α-lactalbumin has a conformation similar to lysozyme that consists of an α-helical lobe and a β-sheet lobe (Brew, 2013).

Another key differentiator between caseins and whey proteins is the phosphorylation that uniquely occurs in caseins, especially at the seryl residues (Farrell et al., 2004). Due to their phosphorylation and hydrophobicity (in the case of β-casein, α_{s1}-casein and α_{s2}-casein), caseins form micellar structures in an aqueous environment (West, 1986), whereas whey proteins are typically present as dissolved monomers. Comparisons of FTIR spectra of intact casein micelles (at ambient pressure) and dissociated casein micelles (at 320MPa pressure) showed no major differences between their absorbance characteristics, except for the appearance of two bands at 990 and 997 cm-1 in fully dissociated casein micelles that correspond to the dianionic phosphate stretch (Gebhardt et al., 2011). The appearance of these bands is related to the release of colloidal calcium from the phosphoseryl residues and an overall increase in negative charge in the casein monomers (Gebhardt et al., 2011). Phosphorylation in caseins makes the proteins' FTIR spectra sensitive to the concentration of calcium. The dianionic phosphate stretch vibrates at a peak of around 976 cm^{-1}, and addition of calcium shifts this peak to higher wavenumbers, while complete protonation of the phosphate groups on the seryl residues at low pH leads to large reductions in the signal from this peak (Fernández et al., 2003).

Partial Least Square Statistical Models. The principles of PLS regressions have been described by Geladi and Kowalski (1986), and the principles of PLS calibration for MIR spectral analysis has been described by Haaland and Thomas (1988). PLS is a useful statistical method for building predictive models when there are many variables involved (e.g. greater than the number of observations) and the variables are highly collinear. In PLS, latent variables are extracted from the independent variables X and dependent variables Y. Latent variables are variables that are not measured directly but are instead inferred from the observed variables based on their detectable effects on the observed variables. This is similar to running a principal component analysis on X, but instead of deriving vectors from X in order to maximize the variance explained in X (minimize the sum of squares of distances of the data points to the vector), PLS aims to derive vectors that maximize the covariance between X and Y, and these new vectors are called "factors" instead of "principal components." In PLS, the X variables are transformed into new variables T, also known as X-scores, and the Y variables are transformed into new variables U, also known as Y-scores, by using loading scores calculated from the vectors. The X and Y scores are chosen such that the covariance of the X and Y variables is maximized.

The PLS factors are ranked based on the amount of variation in the responses that the PLS factors are able to explain, with the first PLS factor explaining the most variation in the response and so on. The first few PLS factors that explain most of the variation in response are used in the predictive model. Cross validation can provide the information to choose the optimal number of factors for a PLS model. Given n spectra samples collected for calibration, cross validation can be done by leaving one

sample out at a time and performing PLS calibrations using n-1 spectra to predict the left out sample until all of the samples have been left out once (Haaland and Thomas, 1988). The predicted concentration of each sample is compared against its reference chemistry values to calculate the prediction error (i.e. SECV), and the least number of factors for which prediction error is minimized are usually chosen as the optimal number of factors.

Developing a PLS model for MIR spectroscopy involves two steps beginning with calibration (model development or training) and followed by validation (prediction testing). For PLS calibration, MIR spectra of a set of samples that have been analyzed by reference chemistry methods for their component concentrations are collected as the population dataset. In PLS calibrations for milk MIR analysis, the independent variables X are MIR absorbance at intervals of wavelengths within the chosen MIR spectral range (e.g. 4000-400cm^{-1}) and the dependent variables Y are the concentration of milk components of interest (e.g. % fat, protein, lactose content). The resulting PLS model from the calibration step is given in the following equation: $Y = \beta_0 + \sum_{i=0}^{k} \beta_i X_i + \epsilon$, where Y is the response variable of milk component concentration, β_0 is the intercept, β_i is the coefficient for the X variables that occur at each wavenumber where absorbance is measured, X_i is the MIR absorbance at each of the spectral wavenumber where it is measured, k is the number of wavenumber points used, and ϵ is the error term. The β coefficients indicate the relationship between the Y response values and each of the X spectral absorbance across the wavenumbers where measurements were taken, and these coefficients provide an indication for which wavenumbers contain spectral information that are most important in predicting the

response (Janik et al., 2007). Following calibration, the PLS model is used to predict the component concentration of a set of independent samples that were not used for calibration and the samples' predicted values are compared with their reference chemistry values for external validation.

Research Objectives

MIR milk analyzers provide rapid and economic determination of the major components in milk, including protein, fat, and lactose. PLS models for the determination of fractions of milk components have been developed for milk fat to predict the composition of individual FAs and groups of FA. Similarly, the same process could be used to create MIR prediction models for the determination of milk protein fractions, which includes CN and SP.

MIR prediction models for the determination of protein in milk are built using calibration sets analyzed using Kjeldahl as a reference chemistry method. Determination of milk CN content using Kjeldahl is based on the amount of milk protein that precipitate at pH 4.6 and is used for the determination of CN%TP in milk. The CN%TP value can serve as a quality indicator for milk and a measure of degree of proteolysis. If the CN content in milk used for determining CN%TP value is defined as only the content of intact CN, the Kjeldahl method may not provide an accurate estimation of CN%TP as milk precipitate at pH 4.6 may also contain CNPP and SP in addition to intact CN. The SDS-PAGE method can quantify the amount of intact CN in milk, which includes α_s-CN, β-CN, and κ-CN, as well as the amounts of CNPP and SP. A comparison of CN%TP results given by Kjedahl and results given by SDS-

PAGE can provide information on the difference in estimation for CN content between the two methods when accounting for intact CN, CNPP, and SP content.

The research objectives were: 1) to develop and validate PLS models using MIR spectra to predict TP, CN, SP, and CN%TP content of micellar casein concentrates and unflavored milk-based beverages; 2) to determine if milk CN%TP estimated by SDS-PAGE is equivalent to CN%TP estimated by Kjeldahl; and 3) to determine the proportion of CN, CNPP, and SP from milk TP that goes into the Kjeldahl NCN filtrate and the proportion that stays in the NCN precipitate using SDS-PAGE.

REFERENCES

AOACI. 2019. Official Methods of Analysis. 21st ed. AOAC International, Gaithersburg, MD.

Andrews, A. T. 1983. Breakdown of caseins by proteinases in bovine milks with high somatic cell counts arising from mastitis or infusion with bacterial endotoxin. J. Dairy Res. 50:57-66.

Banc, A., B. Desbat, and D. Cavagnat. 2011. Ab initio calculations of proline vibrations with and without water: consequences on the infrared spectra of proline-rich proteins. Appl. Spectrosc. 65:817-819.

Barbano, D. M., J. L. Clark, C. E. Dunham, and J. R. Fleming. 1990. Kjeldahl method for determination of total nitrogen content of milk: collaborative study. J. AOAC.73:849-859.

Barbano, D. M., and J. M. Lynch. 1992. Crude and protein nitrogen bases for protein measurement and their impact on current testing accuracy. J. Dairy Sci. 75:3210-3217.

Barth, A. 2000. The infrared absorption of amino acid side chains. Prog. Biophys. Mol. Biol. 74:141–173.

Barth, A. 2007. Infrared spectroscopy of proteins. Biochim. Biophys. Acta. Bioenerg.1767:1073–1101.

Bennett, L. E., W. N. Charman, D. B. Williams, and S. A. Charman. 1994. Analysis of bovine immunoglobulin G by capillary gel electrophoresis. J. Pharm. Biomed. 12: 1103-1108.

Bhattacharjee, C., S. Saha, A. Biswas, M. Kundu, L. Ghosh, and K. P. Das. 2005. Structural changes of β-lactoglobulin during thermal unfolding and refolding – an FT-IR and circular dichroism study. Protein J. 24:27–35.

Boland, M. 2011. Whey Proteins. Pages 30-55 in Handbook of Food Proteins. G.O. Phillips and P.A. Williams, ed. Woodhead Publishing, Cambridge, UK.

Brew, K. 2013. α-Lactalbumin. Pages 261-273 in Advanced Dairy Chemistry: Proteins: Basic Aspects. Vol. 1A. 4th ed. P.L.H. McSweeney and P.F. Fox., ed. Springer US, Boston, MA.

Biggs, D. A. 1967. Milk analysis with the infrared milk analyzer. J. Dairy Sci. 50: 799-803.

Biggs, D. A., G. Johnsson, and L. O. Sjaunja. 1987. Analysis of fat, protein, lactose and total solids by infra-red absorption. Pages 21-30 in Monograph on Rapid

Indirect Methods for Measurement of the Major Components of Milk. Bull. Int. Dairy Fed. No. 208. Int. Dairy Fed., Brussels, Belgium.

Carbonaro, M., and A. Nucara. 2010. Secondary structure of food proteins by Fourier transform spectroscopy in the mid-infrared region. Amino Acids 38:679–690.

Castellino, F. J., W. W. Fish, and K. G. Mann. 1970. Structural studies on bovine lactoferrin. J. Biol. Chem. 245:4269-4275.

Cheng, N., D. M. Barbano, and M. A. Drake. 2019. Effect of pasteurization and fat, protein, casein to serum protein ratio, and milk temperature on milk beverage color and viscosity. J. Dairy Sci. 102:2022–2043.

Creamer, L. K., and T. Richardson. 1984. Anomalous behavior of bovine αs1- and β-caseins on gel electrophoresis in sodium dodecyl sulfate buffers. Arch. Biochem. Biophys. 234:476–486.

Eigel, W. N., J. E. Butler, C. A. Ernstrom, H. M. Farrell, V. R. Harwalkar, R. Jenness, and R. M. Whitney. 1984. Nomenclature of proteins of cow's milk: fifth revision. J. Dairy Sci. 67:1599–1631.

Fagan, C. C., C. P. O'Donnell, L. Rudzik, and E. Wüst. 2009. Milk and Dairy Products. Pages 241-273 in Infrared spectroscopy for food quality analysis and control. 1st ed. D. Sun, ed. Elsevier, Amsterdam, Netherlands.

Farrell, H. M., E. D. Wickham, J. J. Unruh, P. X. Qi, and P. D. Hoagland. 2001. Secondary structural studies of bovine caseins: temperature dependence of β-casein structure as analyzed by circular dichroism and FTIR spectroscopy and correlation with micellization. Food Hydrocoll. 15:341–354.

Fernández, C., S. F. Ausar, R. G. Badini, L. F. Castagna, I. D. Bianco, and D. M. Beltramo. 2003. An FTIR spectroscopy study of the interaction between αs-casein-bound phosphoryl groups and chitosan. Int. Dairy J. 13:897–901.

Fox, P. F., and P. L. H. McSweeney. 2003. Advanced Dairy Chemistry. Springer, New York, N.Y.

Gebhardt, R., N. Takeda, U. Kulozik, and W. Doster. 2011. Structure and stabilizing interactions of casein micelles probed by high-pressure light scattering and FTIR. J. Phys. Chem. B 115:2349–2359.

Geladi, P., and B. R. Kowalski. 1986. Partial least-squares regression: a tutorial. Anal. Chim. Acta 185:1-17.

Goulden, J. D. S. 1964. Analysis of milk by infra-red absorption. J. Dairy Res. 31: 273-284.

Haaland, D. M. and E. V. Thomas. 1988. Partial least-squares methods for spectral analyses. 1. Relation to other quantitative calibration methods and the extraction of qualitative information. Anal. Chem. 60: 1193-1202.

Harwalkar, V. R., H. Cholette, R. C. McKellar, and D. B. Emmons. 1993. Relation between proteolysis and astringent off-Flavor in milk. J. Dairy Sci. 76:2521–2527.

Karman, A. H., and M. A. J. S. van Boekel. 1986. Evaluation of the Kjeldahl factor for conversion of the nitrogen content of milk and milk products to protein content. Neth. Milk Dairy J. 40:315–336.

Kaylegian, K. E., J. M. Lynch, J. R. Fleming, and D. M. Barbano. 2009. Influence of fatty acid chain length and unsaturation on mid-infrared milk analysis. J. Dairy Sci. 92:2485–2501.

Kaylegian, K. E., J. M. Lynch, G. E. Houghton, J. R. Fleming, and D. M. Barbano. 2006. Modified versus producer milk calibration: mid-infrared analyzer performance validation. J. Dairy Sci. 89:2833–2845.

Kelly, P. M., O'Keeffe, A. M., Keogh, M. K., & Phelan, J. A. 1982. Studies of milk composition and its relationship to some processing criteria: III: seasonal variation in heat stability of milk. Irish J. Food Sci. Tech. 6:29-38.

Kindstedt, P. S., A. H. Duthie, and K. M. Nilson. 1983. Estimation of casein from total protein in commingled milk. J. Dairy Sci. 66:2459–2463.

Lynch, J. M., D. M. Barbano, and J. R. Fleming. 1991. Direct and indirect determination of true protein content of milk by Kjeldahl analysis: collaborative study. J. AOAC. 74:281-288.

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., C. Ryan, D. M. Barbano, D. M. Galton, M. A. Rudan, and K. J. Boor. 2000. Effects of Somatic Cell Count on Quality and Shelf-Life of Pasteurized Fluid Milk. J. Dairy Sci. 83:264–274.

Malin, E. L., M. H. Alaimo, E. M. Brown, J. M. Aramini, M. W. Germann, H. M. Farrell, Jr., P. L. H. McSweeney, and P. F. Fox. 2001. Solution structures of

casein peptides: NMR, FTIR, CD, and molecular modeling studies of αs1-Casein, 1–23. J. Protein Chem. 20:391–404.

Manley, M. 2014. Near-infrared spectroscopy and hyperspectral imaging: non-destructive analysis of biological materials. Chem. Soc. Rev. 43:8200–8214.

Maurice-Van Eijndhoven, M. H. T., H. Soyeurt, F. Dehareng, and M. P. L. Calus. 2013. Validation of fatty acid predictions in milk using mid-infrared spectrometry across cattle breeds. Animal 7:348–354.

McParland, S., E. Lewis, E. Kennedy, S. G. Moore, B. McCarthy, M. O'Donovan, S. T. Butler, J. E. Pryce, and D. P. Berry. 2014. Mid-infrared spectrometry of milk as a predictor of energy intake and efficiency in lactating dairy cows. J. Dairy Sci. 97:5863–5871.

Misawa, N., D. M. Barbano, and M. Drake. 2016. Influence of casein as a percentage of true protein and protein level on color and texture of milks containing 1 and 2% fat. J. Dairy Sci. 99:5284–5304.

O'Keeffe, A. M. 1984. Seasonal and lactational influences on moisture content of Cheddar cheese. Irish J. Food Sci. Tech. 8:27-37.

Pasteurized Milk Ordinance. 2017 Revision. Pasteurized Milk Ordinance. The Food and Drug Administration, US Department of Health and Human Services, Washington, DC.

Rienesl, L., N. Khayatzadeh, A. Köck, L. Dale, A. Werner, C. Grelet, N. Gengler, F.-J. Auer, C. Egger-Danner, X. Massart, and J. Sölkner. 2019. Mastitis detection from milk mid-infrared (MIR) spectroscopy in dairy cows. Acta Univ. Agric. Fac. Agron. 67:1221–1226.

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

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

Soyeurt, H., P. Dardenne, F. Dehareng, G. Lognay, D. Veselko, M. Marlier, C. Bertozzi, P. Mayeres, and N. Gengler. 2006. Estimating fatty acid content in cow milk using mid-infrared spectrometry. J. Dairy Sci. 89:3690–3695.

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

Verdi, R. J., D. M. Barbano, M. E. Dellavalle, and G. F. Senyk. 1987. Variability in true protein, casein, nonprotein nitrogen, and proteolysis in high and low somatic cell milks. J. Dairy Sci. 70:230–242.

Wojciechowski, K. L., and D. M. Barbano. 2016. Prediction of fatty acid chain length and unsaturation of milk fat by mid-infrared milk analysis. J. Dairy Sci. 99:8561–8570.

Woolpert, M. E., H. M. Dann, K. W. Cotanch, C. Melilli, L. E. Chase, R. J. Grant, and D. M. Barbano. 2016. Management, nutrition, and lactation performance are related to bulk tank milk de novo fatty acid concentration on northeastern US dairy farms. J. Dairy Sci. 99:8486–8497.

Wrolstad, R. E., T. E. Acree, E. A. Decker, M. H. Penner, D. S. Reid, S. J. Schwartz, C. F. Shoemaker, D. Smith, and P. Sporns. 2005. Characterization of proteins. Pages 155-265 in Handbook of Food Analytical Chemistry. John Wiley & Sons, Inc., Hoboken, New Jersey.

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

CHAPTER TWO

CASEIN MEASURMENT IN MICELLAR CASEIN CONCENTRATE AND MILK PROTEIN BEVERAGES

ABSTRACT

The objective of our study was to develop and validate partial least squares (PLS) models using mid-infrared (MIR) spectra to predict true protein (TP), casein (CN), serum protein (SP), and casein as a percentage of TP (CN%TP) content of micellar casein concentrates (MCC) and unflavored milk-based beverages. The PLS models, developed using 625 samples of formulated milk with a wide range of CN%TP and orthogonal matrix of fat, TP, and lactose were used for model development based on Kjeldahl reference chemistry and MIR spectra collected for each sample. Ingredients for formulation of the 625 samples were produced using cream separation, ultrafiltration, and microfiltration. The ratio of performance deviation (RPD) values were 62.7, 31.4, 14.1, and 3.8 for the prediction of TP, CN, SP, and CN%TP, respectively. External validation of the PLS models using 18 formulated MCC with three different levels of lactose and varying CN%TP resulted in relative standard deviation (RSD) of 0.558, 0.898, 4.257, and 0.591% for prediction of TP, CN, SP, and CN%TP, respectively. The RSD for prediction of CN%TP using a calculated value from TP and CN was lower (0.537 vs 0.591%) than the direct PLS model to predict CN%TP. External validation test of the PLS model using 28 formulated unflavored milk-based beverage with three different levels of fat and varying CN%TP resulted in RSD of 0.350, 1.637, 4.049, and 10.887% for prediction of TP, CN, SP, and CN%TP, respectively.

The RSD for prediction of CN%TP using a calculated value from TP and CN was much lower (1.958 vs 10.89%) than the direct PLS model to predict CN%TP. Some, or all, of the concentrations of TP, CN, and SP in the populations of MCC and milk-based beverage validation samples were outside the range of concentrations of those milk components in the population of samples used to develop the PLS models, yet the validation performance of the PLS models was very good. The ability of these PLS models to perform well on these sample matrices may be due to the high RPD values achieved by orthogonal design of the modeling population used for the PLS model development.

Key Words: casein, mid-infrared, partial least squares

INTRODUCTION

Casein (**CN**) makes up a majority (approximately 82%) of true protein (**TP**) in milk. As such, the CN as a percentage of TP (**CN%TP**) of a milk-based product can greatly influence the product's functionality and serve as a useful indicator of milk quality. Mid-infrared (**MIR**) spectroscopy has the potential to serve as a fast and economic method for the measurement of CN, TP, and CN%TP values in quality control of milk-based product manufacturing. It should be possible to develop a MIR model for the measurement of these values by relating the absorbances arising uniquely from molecular vibrations of different milk proteins with the concentration of those proteins in dairy products.

Differences in the amino acid composition and protein structure of CN and serum proteins (**SP**) may allow for the development of a MIR model to predict the concentration of these two groups of proteins in milk based on differences in their MIR absorbance. Caseins, especially β-CN, are abundant in proline (Holt and Sawyer, 1993). Proline-rich proteins are expected to have a weakened amide II band intensity and a strong band arising from the C-N stretching vibration with a suggested band position around the 1450cm^{-1} region of the FTIR spectra that is largely influenced by water hydration state and hydrogen bonding (Banc et al., 2011). Compared to CN, whey proteins have higher levels of sulfur from cysteine (Fox, 2003). In cysteine, the stretching of the S-H group is shifted to a band position that is free from overlap by other side chain vibrations due to the mass of the sulfur atom (Barth, 2000).

Several studies have applied MIR analysis for the rapid determination of CN in milk. Earlier approaches used filter-based MIR analyzers and determined the CN

content of milk based on the difference in MIR absorbance measurements between milk total protein and non-casein protein from filtrate of the milk precipitated at pH 4.6. Sjaunja and Schaar (1984) used acetic acid to precipitate the CN proteins and corrected the MIR measurements for the non-casein nitrogen (**NCN**) filtrate using partial regression coefficients to adjust for the diluting effect of the added acetic acid and the absorbance of the acid's carboxyl group that often overlaps in the regions of fat and protein absorbance. Barbano and Dellavelle (1987) used phosphoric acid and Karman et al. (1987) used rennet to precipitate CN prior to MIR analysis in order to avoid the challenges associated with acids containing carboxyl groups. These methods require two separate measurements and pretreatment of the milk by precipitation, and thus, still lacked the qualities needed for rapid, real-time testing in dairy production. Later approaches used partial least squares (**PLS**) models to calibrate Fourier transform infrared (**FTIR**) instruments for milk CN measurements without pretreatment of samples. Hewavitharana and van Brakel (1997) built a PLS model for CN prediction, which was calibrated using 270 fresh milk samples, validated using 20 fresh milk samples, and produced a standard error of prediction (**SEP**) and mean difference (**MD**) of 0.066 and 0.041, respectively. Calamari et al. (2009) built a PLS model for CN prediction using 79 calibration and 39 validation samples of fresh milks from 7 dairy farms. The model produced a standard error of cross-validation (**SECV**) and SEP of 0.047 and 0.057, respectively. Another study by Luginbühl (2002) produced four PLS models for CN prediction using FTIR milk analyzers: two models were calibrated using only sets of modified milks produced according to the International Dairy Federation (**IDF**) Standard 141C:2000 and the other two models were calibrated using a

combination of the modified milks and fresh farm milks. The best of the four models, which was calibrated using 38 samples (a combination of modified milks and fresh farm milks), produced a SECV of 0.037 and a SEP and MD of 0.047 and 0.020, respectively (validated using 25 milk samples).

Our study used calibration sample sets of modified milk designed to have an orthogonal matrix design of fat, protein, lactose, and CN%TP concentration. The objective of our study was to develop and validate PLS models using MIR spectra to predict TP, CN, SP, and CN%TP content of micellar casein concentrates (**MCC**) and unflavored milk-based beverages.

MATERIALS AND METHODS

Experimental design

A set of 625 formulated milks varying in fat, TP, CN%TP, and lactose levels was made. All formulated milks were analyzed by MIR and Kjeldahl total nitrogen (**TN**), non-protein nitrogen (**NPN**), and NCN, in duplicate, using Association of Official Analytical Collaboration International (**AOACI**) methods: TN (AOACI, 2019; method 991.20), NPN (AOACI, 2019; method 991.21), and NCN (AOACI, 2019; method 998.05). The MIR spectra plus the data obtained from Kjeldahl analysis of the 625 formulated milks were used to develop PLS models for prediction of TP, CN, SP, and CN%TP in MCC and unflavored milk-based beverages.

For external validation of the PLS models, 2 sets of samples were formulated: 1) a set of 18 MCC varying in lactose and CN%TP levels and 2) a set of 27 unflavored

milk-based beverages varying in fat, TP, CN%TP levels. All samples from both sets were analyzed by MIR and Kjeldahl TN, NPN, and NCN, in duplicate.

Production of 625 formulations

Selection of Target Composition. Fat, TP, CN%TP, and anhydrous lactose content levels were defined to produce samples with a wide range of CN%TP with variation in the background sample matrix (i.e., fat, TP, and lactose) in an orthogonal matrix design. All possible combinations of 4 milk components (fat, TP, CN%TP and anhydrous lactose) with 5 target concentrations for fat (2.5, 3.2, 3.9, 4.6 and 5.3%), 5 for TP (2.5, 3.1, 3.7, 4.3, and 4.9%) within each fat level, 5 for CN%TP (71, 75, 79, 83, and 87%) within each protein level and 5 for lactose (3.5, 4.0, 4.5, 5.0, 5.5%) within each CN%TP level were made, resulting in 625 formulations.

Ingredients. Six different ingredients were used in our formulations: ultrafiltration (**UF**) permeate from skim milk, serum protein isolate (**SPI**), microfiltration (**MF**) retentate, cream, lactose monohydrate (Sigma-Aldrich, PO Box 2060, Milwaukee, WI, and Catalog number L8783), and distilled water. The fat, TP, CN, SP, lactose (anhydrous), and NPN content of each ingredient is shown in Table 2.1. Water was used to extend the lactose range of the samples. UF permeate, SPI, MF retentate and cream were made in Cornell University pilot plant (Ithaca, NY).

Ingredient Production. UF permeate from skim milk was produced as described by Kaylegian et al. (2006). SPI was produced as described by Nelson and Barbano (2005). Cream and MF retentate were produced in the pilot plant at Cornell University. Raw whole milk (930 kg) was pasteurized using a plate heat exchanger system (model 080-S, AGC Engineering, Manassas, VA) at 72°C with a holding time of 16 s and cooled

to 4ºC. Pasteurized milk was heated to 50°C with a plate heat exchanger (model A3, DeLaval Inc., Kansas, MO), and separated using a centrifugal cream separator (model 619, De Laval, Poughkeepsie, NY), to produce 20 kg of cream with about 45.5% fat. The centrifugally separated skim milk was collected and held overnight at ≤4°C. The skim milk was heated to 50°C, prefiltered (NXT 10–30U-M7S, Pall Corporation, Port Washington, NY) into a jacketed stainless steel feed vat (NU-VAT, Meyer-Blanke Company, St. Louis, Missouri), and kept at 50°C while feeding the MF system (Tetra Alcross M7, TetraPak Filtration Systems, Aarhus, Denmark) that contained 0.1 µ ceramic GP Membralox module with channel diameter of 4 mm and membrane area of 1.68 m^2 (EP1940GL0.1µA, Alumina, Pall Corporation). The MF system was operated as described by Hurt et al. (2010).

In the first MF stage, skim milk (296 kg) containing 3.10% TP was processed to a 3.41X concentration factor (**CF**) (based on weight of permeate removed) at 50°C to produce retentate with a TP concentration of 9.2%. The typical inlet pressures were 435 kPa for retentate and 250 kPa for permeate, and typical outlet pressures were 240 kPa for retentate and 260 kPa for permeate. Permeate mass flux (kg/m²/h), CF, and system conditions were measured every 15 min, and samples of permeate and retentate were taken for MIR analysis (LactoScope FTIR, PerkinElmer Analytical Solutions formerly Delta Instruments, B.V., Kelvinlaan 3. 9207 JB Drachten, The Netherlands). At the end of the stage, the collected retentate (83 kg) and the collected permeate (200 kg) were mixed separately, sampled, and samples were cooled to <4°C. One sample of each was taken for MIR analysis and the other samples were frozen at -80°C to be analyzed later

using the Kjeldahl method. Permeate was discarded and retentate was used as feed for next stage.

In the second stage, retentate (83 kg) from first stage was mixed with reverse osmosis (**RO**) water used for diafiltration (200 kg at a diafiltration factor of 3.66), sampled for MIR analysis, and recirculated for 15 min. The valves were adjusted to attain the retentate and permeate removal rates of 45.6 and 125.0 L/h, respectively. The retentate recirculation rate was set to approximately 706 L/min with a cross-flow velocity of 7 m/s. The feed containing about 2.63% TP, was processed to a 3.66X CF at 50°C to produce retentate with TP concentration of 9.2 %. The typical inlet pressures were 424 kPa for retentate and 226 kPa for permeate, and typical outlet pressures were 233 kPa for retentate and 240 kPa for permeate. Permeate mass flux ($kg/m^2/h$), CF, and system conditions were measured every 15 min, and samples of permeate and retentate were taken for MIR analysis. At the end of the second stage, the collected retentate (73.50 kg) and the collected permeate (195 kg) were mixed separately, sampled, and samples were cooled to <4°C. One sample of each was taken for MIR analysis and the other samples were frozen at -80°C to be analyzed later by Kjeldahl. Permeate was discarded and retentate was used as feed for next stage.

In the third stage, retentate (73.50 kg) from second stage was mixed with RO water used for diafiltration (195.50 kg at a diafiltration factor of 3.43), sampled for MIR analysis, and recirculated for 15 min. The valves were adjusted to attain the retentate and permeate removal rates of 60 and 156 L/h, respectively. The retentate recirculation rate was set to approximately 702 L/min with a cross-flow velocity of 7 m/s. The feed containing 2.35% TP, was processed to a 3.43X CF at 50°C to produce retentate (i.e.,

95.6% SP reduced MCC with our target TP concentration of 8.728%. The typical inlet pressures were 418 kPa for retentate and 208 kPa for permeate, and typical outlet pressures were 233 kPa for retentate and 221 kPa for permeate. Permeate mass flux (kg/m²/h), CF, and system conditions were measured every 15 min, and samples of permeate and retentate were taken for MIR analysis. At the end of the stage, the collected retentate (70 kg) and the collected permeate (169 kg) were mixed separately, sampled, and samples were cooled to <4°C. One sample of each was taken for MIR analysis and the other samples were frozen at -80°C to be analyzed later by Kjeldahl. Permeate was discarded and retentate was split into 4L plastic containers with screw caps and refrigerated at <4°C. The after-processing cleaning was operated as described by Hurt et al. (2010).

Calculation of Weights of Ingredients. A Microsoft Excel (Microsoft, Corp., Seattle, WA) Solver add-in was set up to calculate the target weight of each of the 6 ingredients, in grams, that was necessary to reach the TP, CN%TP, fat, and lactose concentration target for each formulation. The calculations were based on the ingredient compositions shown on Table 2.1.

Table 2.1. Composition (%) of fat, true protein (TP), casein (CN), serum protein (SP), anhydrous lactose, and non-protein nitrogen (NPN) within each ingredient used to formulate the 625 milks.

	Fat	TP	CN	SP	Lactose	NPN
UF Permeate	0.056	0.559	0.000	0.559	4.725	0.170
SPI	0.000	24.483	1.354	23.128	0.100	0.224
MCC	0.173	8.728	8.342	0.386	0.540	0.053
Cream	45.500	1.540	1.240	0.300	2.720	0.11

Lactose monohydrate	0.00	0.00	0.00	0.00	95.00	0.00
Water	0.00	0.00	0.00	0.00	0.00	0.00

UF = ultrafiltration, SPI = serum protein isolate, MCC = micellar casein concentrate

Formulations. The ingredients used for formulation were weighed using three analytical balances with different range of readabilities (**d**), and capacities depending on the weight of the ingredient need in the formulation. Lactose and water were weighed first, in a balance that had a maximum capacity of 310g and d = 0.1mg (model AX304, Mettler Toledo, Switzerland). Next, UF permeate and MF retentate were weighed in a balance that had a maximum capacity of 5100g and d = 0.01g (model PG5002-S, Mettler Toledo, Switzerland). Lastly, SPI and cream were weighed using a balance that had a maximum capacity of 405g and d = 0.1mg (model AT400, Mettler Toledo, Switzerland). The total target batch weight of each of the 625 formulations was 250 grams. Next, each formulation was mixed, split between one 90 mL vial (Capitol Plastic Products, 1030 Riverfront Center, Amsterdam, NY, and Catalog number CPP03EDM-CL), which was immediately analyzed by MIR, and three 60 mL vials (Capitol Plastic Products, 1030 Riverfront Center, Amsterdam, NY, and Catalog number CPP02CL), which were frozen using liquid nitrogen and kept at -80°C.

Mid-Infrared Analysis. Samples collected during MF run were analyzed by MIR to monitor the composition and control the process. In addition, each one of the 625 formulations were analyzed by MIR to monitor the composition and generate the spectra that were used for PLS modeling. MIR analysis was performed with the Delta LactoScope FTIR Advanced (model FTA) milk analyzer (PerkinElmer Analytical Solutions, Drachten, The Netherlands) equipped with a BMX optical bench (ABB

Bomem, Montreal, Canada). A flow through CaF_2 cuvette (36 µm) was used. For collection of the MIR spectra, the LactoScope FTA was operated at a spectral resolution of 8 cm^{-1} (digital resolution of 3.85 cm^{-1}) collecting averages of 16 scans per analysis (using standard double-sided forward-backward interferogram acquisition, cosine apodization and phase correction for Fourier transformation yielding an energy spectrum). The average of 16 scans was transformed into an absorption spectrum by calculation of the -log ratio with reference to the spectrum of zero liquid (demineralized water containing 0.01% triton) and storage of the absorption spectrum for each sample using FTIRscope Advanced software (PerkinElmer Analytical Solutions). The range of wave numbers in the spectra was from 400 cm^{-1} to 4000 cm^{-1}.

Kjeldahl Analysis. Three days before the Kjeldahl analysis, 25 samples were transferred from a -80°C to a -20°C freezer. There were 2 vials of each frozen formulated milk and one was used for NCN analysis on day 1 and the other for TN plus NPN analysis on day 2. On day 1, samples were thawed using a microwave (Model 721.69072901, Sears, Roebuck and Co., Hoffman Estates, IL) set to bring sample temperature to 4°C. Formulated milks were prepared for NCN (AOACI, 2019; method 998.05) in duplicate. The acetic acid and sodium acetate volumes used were determined as recommended by Wojciechowski and Barbano (2015). On day 2, samples were thawed using a microwave and prepared for TN (AOACI, 2019; method 991.20) and NPN (AOACI, 2019; method 991.21) in duplicate. TP was calculated by subtracting NPN from TN and multiplying by the factor 6.38. CN was calculated by subtracting NCN from TN and multiplying by the factor 6.38.

Development of the PLS Statistical Models

Removal of Outliers. Of the total 625 formulated milks, the data set for TP, CN, SP, and CN%TP model development had 37, 29, 46, and 23 samples removed, respectively, (which were spectral outliers identified in the PLS modeling processes) based on the Mahalanobis distance.

Calculation of PLS Models. The PLS models for prediction of TP, CN, SP, and CN%TP were calculated using the following spectral ranges: 3000 to 2750, 1800 to 1700 and 1585 to 1000 cm^{-1}. Data were mean centered**.** No variance scaling or baseline correction was used. A one sample out cross validation was applied for determining the number of factors and SECV for each model. The number of factors selected for each final prediction model was the number of factors that produced an F-Test predicted residual sum of squares (**FPRESS**) between 0.5 and 1, which will be near the minimum SECV. The beta coefficients for the PLS prediction models were calculated by Grams A/I PLSIQ Version 7.00 software (Thermo Fisher Scientific Inc., Waltham, MA), using MIR spectra and Kjeldahl results for each formulated milk. The PLS prediction models are using the following equation: $Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k + e$, where Y = response [i.e., TP, CN, SP, and CN%TP values], X = MIR spectral absorbance, β = beta coefficients, and e = error term. The PLS models were installed into the FTIRScope Advanced software.

External Performance Validation of the PLS Models

Formulation of Micellar Casein Concentrates for External PLS Model Validation. The purpose of this evaluation of the models was to determine their

performance when applied to a population of sample that might be tested in milk processing factor to determine the degree of removal of SP from skim milk during a MF process. Fat, anhydrous lactose, TP, and CN%TP formulation target concentrations were defined to produce MCC with 3 levels of lactose and a wide range of CN%TP (Table 2.2). Five ingredients were used in our formulations: SPI, MCC stage 1, MCC stage 2, MCC stage 3, and distilled water. SPI was produced as described by Nelson and Barbano (2005). The MCC were produced at North Carolina State University as described by Hurt et al. (2010).

Table 2.2. Target composition (%) of fat, anhydrous lactose, true protein (TP), and casein as percentage of true protein (CN%TP), within the validation set containing 18 formulated micellar casein concentrates (MCC).

MCC	Fat	Lactose	TP	CN%TP
1	0.4	4.2	8.6	80
2	0.4	4.2	8.6	82
3	0.4	4.2	8.6	84
4	0.4	4.2	8.6	86
5	0.4	4.2	8.6	88
6	0.4	4.2	8.6	91
7	0.4	1.5	8.6	84
8	0.4	1.5	8.6	86
9	0.4	1.5	8.6	88
10	0.4	1.5	8.6	90
11	0.4	1.5	8.6	92
12	0.4	1.5	8.6	94
13	0.4	0.9	8.6	86
14	0.4	0.9	8.6	88
15	0.4	0.9	8.6	90

16	0.4	0.9	8.6	92
17	0.4	0.9	8.6	94
18	0.4	0.9	8.6	96

A Microsoft Excel (Microsoft, Corp., Seattle, WA) solver spreadsheet was used to calculate the weight of each of the 5 ingredients, in grams, that was necessary to reach the fat, lactose, TP and CN%TP concentration target for each formulation, as described above from making the 625 formulations used in development of the PLS models. The total target batch weight of each of the 18 formulations was 360 grams. Next, each formulation was mixed, split among one 90 mL vial (Capitol Plastic Products, 1030 Riverfront Center, Amsterdam, NY, and Catalog number CPP03EDM-CL), which was immediately analyzed by MIR, and three 60 mL vials (Capitol Plastic Products, 1030 Riverfront Center, Amsterdam, NY, and Catalog number CPP02CL), which were frozen using liquid nitrogen and stored at -80°C. Each formulated MCC was analyzed by Kjeldahl for TN, NPN, and NCN, in duplicate, as described above.

Formulation of Unflavored Milk-based Beverages for External PLS Model Validation. Fat, anhydrous lactose, TP, and CN%TP content levels were selected to produce a population of unflavored milk-based beverages with 3 levels of fat, 3 levels of TP, and a wide range of CN%TP that enabled a validation of the TP, CN, and SP PLS model performance when applied for analysis of milk based beverages that contained about 15 g of protein per 240 g serving (Table 2.3). Five ingredients were used in our formulations: cream, anhydrous lactose (Sigma-Aldrich, PO Box 2060, Milwaukee, WI, and Catalog number L8783), MCC stage 3, SPI, and distilled water. Cream was made in Cornell University pilot plant (Ithaca, NY). The MCC was produced at North

Carolina State University as described by Hurt et al. (2010). SPI was produced as described by Nelson and Barbano (2005).

Table 2.3. Target composition (%) of fat, anhydrous lactose, true protein (TP), and casein as percentage of true protein (CN%TP), within the validation set containing 27 formulated unflavored milk-based beverages.

Unflavored beverage	Fat	Lactose	TP	CN%TP
1	0.2	2.2	3.0	40
2	0.2	2.2	3.0	60
3	0.2	2.2	3.0	80
4	1.0	2.2	3.0	40
5	1.0	2.2	3.0	60
6	1.0	2.2	3.0	80
7	2.0	2.2	3.0	40
8	2.0	2.2	3.0	60
9	2.0	2.2	3.0	80
10	0.2	2.2	6.6	40
11	0.2	2.2	6.6	65
12	0.2	2.2	6.6	90
13	1.0	2.2	6.6	40
14	1.0	2.2	6.6	65
15	1.0	2.2	6.6	90
16	2.0	2.2	6.6	40
17	2.0	2.2	6.6	65
18	2.0	2.2	6.6	90
19	0.2	2.2	8.5	40
20	0.2	2.2	8.5	70
21	0.2	2.2	8.5	95
22	1.0	2.2	8.5	40
23	1.0	2.2	8.5	70
24	1.0	2.2	8.5	95
25	2.0	2.2	8.5	40
26	2.0	2.2	8.5	70
27	2.0	2.2	8.5	95

A Microsoft Excel (Microsoft, Corp., Seattle, WA) solver spreadsheet was used to calculate the weight of each of the 5 ingredients, in grams, that were necessary to

reach the fat, lactose, TP and CN%TP concentration targets for each formulation that are shown in Table 2.3. The total target batch weight of each of the 27 formulations was 360 grams. Next, each formulation was mixed, split among one 90 mL vial (Capitol Plastic Products, 1030 Riverfront Center, Amsterdam, NY, and Catalog number CPP03EDM-CL), which was immediately analyzed by MIR, and three 60 mL vials (Capitol Plastic Products, 1030 Riverfront Center, Amsterdam, NY, and Catalog number CPP02CL), which were frozen using liquid nitrogen and kept at -80°C. Each formulated unflavored milk-based beverage was analyzed by Kjeldahl TN, NPN, and NCN, in duplicate, as described above.

Validation of the PLS Models. The PLS models were slope and intercept adjusted separately using the MIR predictions of the TP, CN, SP, and CN%TP content of the MCC and unflavored milk-based beverage validation sets, and using the Kjeldahl TN, NPN, and NCN reference chemistry for each validation set. A linear regression equation for each predicted parameter was calculated by plotting reference values given by Kjeldahl as a function of MIR predicted values (X-axis). These linear regression equations provided the slope and intercept adjustment of each predicted parameter to achieve a mean difference between reference values and MIR predictions of zero. This allowed for the determination of the residual unexplained variation in the results (i.e., standard deviation of difference) produced by the PLS models when testing the external validation samples.

RESULTS AND DISCUSSION

Development of the PLS Statistical Models

The number of samples, mean of Kjeldahl reference values, standard deviation (**SD**), and minimum and maximum values used for development of each model are shown in Table 2.4. Our goal was to develop robust models that would work over a range of different milk matrices. A 6 factor model was selected as the optimum model for prediction of TP and an 8 factor model was selected as the optimum model for prediction of CN, SP, CN%TP. The ratio of performance deviation (**RPD**), which is the standard deviation of final population of reference data (i.e., Kjeldahl) divided by SECV, is a parameter used to indicate how robust is each developed model (Di Marzo et al., 2016). According to Manley M. (2014), a model with RPD higher than 3 can be used for screening, higher than 5 can be used for quality control, and higher than 8 can be used for any application. The PLS models for prediction of TP, CN, and SP were excellent, with RPD of 62.7, 31.4, and 14.1, respectively. The PLS model for prediction of CN%TP had a lower RPD of 3.8 (Table 2.4). These four PLS models for prediction of TP, CN, SP, and CN%TP were assigned PerkinElmer Analytical Solutions parameter numbers 7401, 7402, 7403, and 7404, respectively, were compiled and run on Delta Instruments MIR milk analyzers.

The β-coefficients for the two PLS models for direct prediction of milk casein and serum protein were compared to determine where differences in magnitude and sign of the β-coefficients differed. The wavelengths where key differences in β-coefficients occurred were in the area of the MIR spectra where the amino acid proline absorbs (Barth 2000, 2007).

Table 2.4. Number of samples, mean of Kjeldahl reference value, standard deviation (SD), minimum and maximum values of final population of samples (i.e., with outliers removed) used to develop mid-infrared (MIR) partial least squares (PLS) prediction models[1]

	TP	CN	SP	CN%TP
Number of samples	588	596	579	602
Mean	3.726	2.867	0.848	76.962
SD	0.869	0.692	0.258	4.702
Minimum	2.467	1.746	0.377	68.440
Maximum	5.043	4.192	1.524	85.020
Number of factors	6	8	8	8
SECV	0.014	0.022	0.018	1.241
R-square	1.000	0.999	0.995	0.930
F-Ratio (PRESS)	1.056	1.163	1.254	1.007
F-Test (FPRESS)	0.746	0.968	0.997	0.534
RPD	62.7	31.4	14.1	3.8

[1]Number of factors, standard error of cross validation (SECV), R-square, F-ratio predicted residual sum of squares (PRESS), F-test (FPRESS), and the ratio of standard deviation to standard error of cross validation (RPD) of the calibration step for the PLS model to predict true protein (TP (%)), casein (CN (%)), serum protein (SP (%)), and casein as percentage of true protein (CN%TP).

Validation Performance of PLS Models When Testing MCC

The MD, standard deviation of the difference (**SDD**), and relative standard deviation (**RSD**) of the difference (i.e., coefficient of variation) were calculated for TP, SP, CN, and CN%TP for the MCC external validation sample set (Table 2.5). The models for prediction of TP, CN, SP, and CN%TP resulted in RSD of 0.558, 0.898, 4.257, and 0.591%, respectively. A better RSD (0.537%) was achieved for CN%TP by calculating it from MIR predictions of CN and TP, than by predicting CN%TP directly with a PLS model (Table 2.5). This is due to the much higher RPD values (Table 2.4) for the CN and TP models compared to the RPD for the direct CN%TP model. Using the separate results of the more robust CN and TP models to calculate CN%TP gave

slightly better validation performance on MCC than the direct prediction of CN%TP. The SP PLS model had a SDD of 0.046%, which was the same as the SDD for the TP model, but because the mean concentration of SP is lower than the mean concentration of TP, this resulted in a RSD of 4.257%. According to Wojciechowski and Barbano (2016), higher RSD are obtained in general when the concentration of the measured component is lower. It is interesting to note that the concentration of TP, CN, and SP in the MCC validation set (Table 2.5) are much higher than the concentration of the same components in the population of samples used for development of the PLS models (Table 2.4), yet the PLS models performed very well when testing MCC at over 9% TP (Table 2.5). This indicates that the models are very robust and this is consistent with their high RPD values (Table 2.4).

The RSD of the difference provides a measure of the variability of the model's prediction across a range of component concentrations when measured against reference chemistry values. According to AOACI (Appendix F, Table A4) (2019), the expected RSD of a method as a function of analyte concentrations are 3.7, 2.7, 1.9, and 1.3% for analyte concentrations of 0.1%, 1%, 10%, and 100%, respectively. Under these conditions, the CN and TP models had excellent repeatability with RSD < 1% for analyte concentrations of less than 10%, while the SP model had poorer RSD than typically expected for analyte concentrations in the range of 0.5 to 2%.

Table 2.5. Number of samples, mean of Kjeldahl reference values, mean of mid-infrared (MIR) predicted values, mean difference (MD) between MIR predicted and Kjeldahl reference values, range of Kjeldahl reference values, range of MIR predicted values, standard deviation of the difference (SDD), and relative standard deviation (RSD) of micellar casein concentrates (MCC) used for validation of PLS models to predict true protein (TP), casein (CN), serum protein (SP), and casein as percentage of true protein (CN%TP).

	TP (%)	CN (%)	SP (%)	CN%TP	CN%TP (calculated)[1]
Number of Samples	18	18	18	18	18
Kjeldahl (reference) Mean	9.401	8.324	1.077	88.54	88.54
MIR (predicted) Mean	9.401	8.324	1.077	88.54	88.54
MD	0	0	0	0	0
Range Kjeldahl	0.494	1.129	1.029	11.02	11.02
Range MIR	0.408	1.105	0.974	10.65	10.47
SDD	0.052	0.075	0.046	0.524	0.476
RSD	0.558	0.898	4.257	0.591	0.537

[1]CN%TP (calculated) = (MIR predicted CN/MIR predicted TP) * 100

Validation Performance of PLS Models When Testing Unflavored Milk-based Beverage

The MD, SDD, and RSD of the difference were calculated for TP, SP, CN, and CN%TP for the unflavored milk-based beverage external validation sample set (Table 2.6). The models for prediction of TP, CN, SP, and CN%TP resulted in RSD of 0.350, 1.637, 4.049, and 10.887%, respectively. The RSD achieved for CN%TP by calculating it from MIR predictions of CN and TP (1.958%) was over five times better than by predicting CN%TP directly with a PLS model (Table 2.6) because the RPD values (Table 2.4) for the CN and TP models are much higher than the direct CN%TP model. Considering the range of expected RSD values described by AOACI (2019), the TP and

CN prediction models produced RSD within expected values for the concentration range used in the validation. Again, it is interesting to note that the mean concentration of TP, CN, and SP in the milk-based beverage validation set (Table 2.6) are higher and the range of CN%TP is larger than the values for these parameters in the population of samples used for development of the PLS models (Table 2.4), yet the PLS models performed very well when testing milk-based beverages at over 6% TP (Table 2.6). This indicates that the models are very robust and this is consistent with their high RPD values (Table 2.4).

Table 2.6. Number of samples, mean of Kjeldahl reference values, mean of mid-infrared (MIR) predicted values, mean difference (MD) between MIR predicted and Kjeldahl reference values, range of Kjeldahl reference values, range of MIR predicted values, standard deviation of the difference (SDD), and relative standard deviation (RSD) of unflavored milk-based beverages used for validation of PLS models to predict true protein (TP), casein (CN), serum protein (SP), casein as percentage of true protein (CN%TP).

	TP (%)	CN (%)	SP (%)	CN%TP	CN%TP (calculated)[1]
Number of Samples	27	27	27	27	27
Kjeldahl (reference) Mean	6.370	4.398	1.972	67.35	67.35
MIR (predicted) Mean	6.370	4.398	1.972	67.35	67.35
MD	0	0	0	0	0
Range Kjeldahl	6.265	7.288	3.956	47.43	47.43
Range MIR	6.287	7.325	3.983	52.09	46.52
SDD	0.022	0.072	0.080	7.332	1.319
RSD	0.350	1.637	4.049	10.887	1.958

[1]CN%TP (calculated) = (MIR predicted CN/MIR predicted TP) * 100

CONCLUSIONS

The PLS models, developed using 625 samples of formulated milk with a range of CN%TP and orthogonal matrix of fat, TP, and lactose, yielded RPD values of 62.7, 31.4, 14.1, and 3.8 for the prediction of TP, CN, SP, and CN%TP, respectively. External validation of the PLS models using 18 formulated MCC with three different levels of lactose and varying CN%TP resulted in RSD of 0.558, 0.898, 4.257, and 0.591% for prediction of TP, CN, SP, and CN%TP, respectively. The RSD for prediction of CN%TP in the MCC validation sample set using a calculated value from TP and CN was lower (0.537 vs 0.591%) than the direct PLS model to predict CN%TP. External validation test of the PLS model using 28 formulated unflavored milk-based beverage with three different levels of fat and varying CN%TP resulted in RSD of 0.350, 1.637, 4.049, and 10.887% for prediction of TP, CN, SP, and CN%TP, respectively. The RSD for prediction of CN%TP in the unflavored milk-based beverage validation sample set using a calculated value from TP and CN was much lower (1.958 vs 10.89%) than the direct PLS model to predict CN %TP. Some, or all, of the concentrations of TP, CN, and SP in the populations of MCC and milk-based beverage validation samples were outside the range of concentrations of those milk components in the population of samples used to develop the PLS models, yet the validation performance of the PLS models was very good. The ability of these PLS models to perform well on these sample matrices may be due to the high RPD values achieved by orthogonal design of the modeling population used for the PLS model development.

ACKNOWLEDGEMENTS

This research was supported in part by the UDSA Federal Milk Market orders and PerkinElmer Analytical Solutions, Drachten, The Netherlands. The technical assistance of Michelle Bilotta, Chassidy Coon, Sara Hasting with reference chemistry analysis, Larrisa Di Marzo with formulation of the orthogonal sample set, and Ni Cheng for production of micellar casein concentrate was greatly appreciated.

REFERENCES

AOACI. 2019. Official Methods of Analysis. 21st ed. AOAC International, Gaithersburg, MD.

Banc, A., B. Desbat, and D. Cavagnat. 2011. Ab initio calculations of proline vibrations with and without water: consequences on the infrared spectra of proline-rich proteins. Appl. Spectrosc. 65: 817-819.

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.

Barth, A. 2000. The infrared absorption of amino acid side chains. Prog. Biophys. Mol. Biol. 74:141–173.

Barth, A. 2007. Infrared spectroscopy of proteins. Biochim. Biophys. Acta. Bioenerg.1767:1073–1101.

Calamari, L., M. G. Maianti, S. Ferrandi, P. Franceschi, E. Cremaschi, P. Formaggioni, and A. Summer. 2009. Applicazione della spettroscopia FT-IR per la determinazione della caseina in campioni individuali di latte bovino. Scienza E Tecnica Lattiero-Casearia 60: 355-359.

Di Marzo, L., P. Cree, and D. M Barbano. 2016. Prediction of fat globule particle size in homogenized milk using Fourier transform mid-infrared spectra. J. Dairy Sci. 99:8549-8560.

Fox, P.F. 2003. Milk proteins: general and historical aspects. Page 10 in Advanced Dairy Chemistry – 1 Proteins. 3rd ed. P. F. Fox and P.L.H. McSweeney, ed. Springer, New York, NY.

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.

Holt, C. and L. Sawyer. 1993. Caseins as rheomorphic proteins: interpretation of the primary and secondary structures of the α_{s1}-, β-, and κ-caseins. J. Chem. Soc. Faraday Trans., 89:2683-92.

Hurt, E., J. Zulewska, M. Newbold, and D. M. Barbano. 2010. Micellar casein concentrate production with a 3X, 3-stage, uniform transmembrane pressure ceramic membrane process at 50°C. J. Dairy Sci. 93: 5588-5600.

Kaylegian, K. E., G. E. Houghton, J. M. Lynch, J. R. Fleming, and D. M. Barbano. 2006. Calibration of infrared milk analyzers: modified milk versus producer milk. J. Dairy Sci. 89:2817–2832.

Karman, A. H., M. S. van Boekel, and A. P. Arentsen-Stasse 1987. A simple and rapid method to determine the casein content of milk by infrared spectrophotometry. Neth. Milk Dairy J. 41: 175-187.

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.

Manley, M. Near-infrared spectroscopy and hyperspectral imaging: non-destructive analysis of biological materials. Chem. Soc. Rev. 43:8200-8214.

Nelson, B. K., and D. M Barbano. 2005. A Microfiltration process to maximize removal of serum proteins from skim milk before cheese making. J. Dairy Sci. 88:1891-1900.

Sjaunja, L. O., and J. Schaar. 1984. Determination of casein in milk by infrared spectrophotometry. Milchwissenschaft 39: 288-290.

Wojciechowski, K. L. and D. M. Barbano. 2015. Modification of the Kjeldahl noncasein nitrogen method for bovine milk concentrates and milks from other species. J. Dairy Sci. 98: 7510-7526.

Wojciechowski, K. L. and D. M. Barbano. 2016. Prediction of fatty acid chain length and unsaturation of milk fat by mid-infrared milk analysis. J. Dairy Sci. 99:8561-8570.

CHAPTER THREE

MEASUREMENT OF CASEIN IN MILK BY KJELDAHL AND SODIUM DODECYL SULFATE POLYACRYLAMIDE GEL ELECTROPHORESIS

ABSTRACT

Our objectives were to determine if milk casein as percentage of true protein (CN%TP) estimated by sodium dodecyl sulfate polyacrylamide gel (SDS-PAGE) is equivalent to CN%TP estimated by Kjeldahl and to determine the proportion of casein (CN), casein proteolysis products (CNPP), and serum protein (SP) from milk true protein (TP) that goes into the Kjeldahl non-casein nitrogen (NCN) filtrate and the proportion that stays in the NCN precipitate using SDS-PAGE. Raw milk samples were collected from 16 mid-lactation Holstein cows twice a week for two weeks. These milks were analyzed for Kjeldahl total nitrogen (TN), non-protein nitrogen (NPN), and NCN content in duplicate, and by SDS-PAGE. The CN%TP determined by Kjeldahl was compared with the CN%TP estimated by SDS-PAGE calculated in two ways: as a percentage of only intact caseins divided by true protein and as a percentage of both intact caseins and CNPP divided by true protein. Three milks varying in fat, lactose, TP, CN, and SP content were formulated. These milks were analyzed in duplicate for Kjeldahl TN, NPN, and NCN content, and each of the NCN filtrate and NCN precipitate were analyzed in duplicate by SDS-PAGE for relative quantity (%) of CN, CNPP, and SP. We found that the estimate of CN%TP by Kjeldahl was higher than the estimate of CN%TP by SDS-PAGE that was calculated as only intact CN divided by the total of all protein bands. However, no difference was detected in the estimate of CN%TP by

Kjeldahl compared to CN%TP by SDS-PAGE when CNPP were included as CN in the calculation of SDS-PAGE results. Based on SDS-PAGE results, we found that a majority (89%) of the CNPP from the milk (approximately 10.13% out of 11.41% TP) were retained in the Kjeldahl NCN precipitate. Thus, CN%TP measured by Kjeldahl underestimates the amount of proteolytic damage that has been done to casein in milk.

Key Words: Casein, casein proteolysis, Kjeldahl, SDS-PAGE

INTRODUCTION

In the US dairy market, the milk price calculation places strong economic emphasis on milk protein and fat (USDA, 2017). The true protein (**TP**) portion of milk (about 80 to 82% casein (**CN**) has a significant impact on cheese yield (Margolies et al., 2017) and nutritional value (essential amino acids and carrier of calcium) of fluid milk and dairy products (Fox et al., 2015).

The Kjeldahl method is the reference method for milk protein determination (AOAC, 2000) based on the nitrogen content of milk and milk fractions. This method assumes that on average all milk proteins contain about 15.67% nitrogen and uses the resulting factor of 6.38 to convert total nitrogen (**TN**) to total protein (Karman and Van Boekel, 1986). The determination of nitrogen using the Kjeldahl method involves three steps: digestion to release nitrogen from protein, distillation to collect nitrogen in boric acid as ammonia, and titration to quantify the amount of nitrogen collected as a percentage by weight of the sample *(*Barbano et al., 1990; Lynch and Barbano, 1999).

In 1938, Samuel J. Rowland demonstrated that nitrogen in milk was distributed among several groups of proteins and that a portion of the nitrogen content in milk was

not associated with proteins (Rowland,1938a,b). There are three primary nitrogen fractions of milk that are commonly measured with the Kjeldahl method today: TN, non-protein nitrogen (**NPN**), and non-casein nitrogen (**NCN**) based on the Rowland milk nitrogen fraction scheme. The NPN is the nitrogen soluble in 12% TCA and the NCN is the nitrogen soluble in a pH 4.6 acetate buffer solution determined as described in official methods (AOAC, 2000). The TN minus the NPN multiplied by 6.38 is the TP content of milk. The TN minus NCN multiplied by 6.38 is defined as CN. The TP has been used as the basis for milk protein payment by the USDA federal milk markets since January 2000. Using TP as the calibration basis for protein testing using mid-infrared (**MIR**) analysis improved the accuracy of milk protein testing (Barbano and Lynch, 1992). The CN as percentage of TP (**CN%TP**) has been used as an index of milk protein quality because CN%TP decreases as proteolytic enzymes breakdown CN. However, the Kjeldahl NCN method assumes that filtrate contains only the NCN components of milk. Karman and Van Boekel (1986) suggested that during the Kjeldahl NCN analysis, a portion of the CN proteolysis products **(CNPP)** remain with the NCN precipitate and another portion of the CNPP go into the NCN filtrate. Detailed method descriptions and collaborative studies of each of these methods have been published (Barbano et al., 1990, 1991; Lynch et al., 1998; Wojciechowski and Barbano, 2015).

The sodium dodecyl sulfate polyacrylamide gel (**SDS-PAGE**) is a secondary method that has been used to quantify the relative proportions of CN, CNPP, and serum proteins (**SP**) in milk (Verdi et al., 1987). In the SDS-PAGE method, a portion of milk is dissolved in a denaturing sample buffer solution that contains SDS anions and dithiothreitol (**DTT**). The SDS anions attach onto the protein, forming SDS-protein

complexes that have similar and high negative charge density (Wrolstad et al., 2005). The milk dissolved in sample buffer is loaded on to a polyacrylamide gel and an electric field is applied to the system. The milk proteins are separated based on their molecular weight. At the end of the run, the protein bands are stained with a dye, the gel is scanned, the bands are quantified and the trace quantity, expressed as mm x optical density (**OD**), of each protein band is given as result (Wrolstad et al., 2005).

In the literature, there is no specific information comparing the results given by Kjeldahl versus the results given by SDS-PAGE for estimation of CN%TP on the same milks. Our objectives were to determine if milk CN%TP estimated by SDS-PAGE is equivalent to CN%TP estimated by Kjeldahl and to determine the proportion of CN, CNPP, and SP from milk TP that goes into the Kjeldahl NCN filtrate and the proportion that stays in the NCN precipitate using SDS-PAGE.

MATERIALS AND METHODS

Experimental Design

Raw milks from 16 mid-lactation Holstein cows were collected twice in each of two weeks for a total of 64 samples. All milks were analyzed for Kjeldahl TN, NPN, and NCN content in duplicate, and by SDS-PAGE. The CN%TP given by Kjeldahl was calculated and compared with the CN%TP estimated by SDS-PAGE calculated in two ways: 1) using only intact CN bands divided by the total of all protein bands (i.e., CN+CNPP+SP) and 2) including the sum of intact CN and CNPP bands divided by the total of all protein bands.

Next, three milks varying in fat, lactose, TP, CN, and SP content were formulated as described by Kaylegian et al. (2006) and analyzed for Kjeldahl TN, NPN, and NCN content, in duplicate. During the Kjeldahl NCN sample preparation step, each NCN filtrate and NCN precipitate was collected and analyzed by SDS-PAGE, in duplicate. The relative quantity (%) of CN, CNPP, and SP in each NCN filtrate and NCN precipitate was estimated and the proportion (%) of CN, CNPP, and SP from formulated milk that goes into the Kjeldahl NCN filtrate and the proportion that stays in the NCN precipitate was determined.

Determination of CN%TP by Kjeldahl versus CN%TP by SDS-PAGE

Kjeldahl Analysis. Each of the 64 individual cow milks, was analyzed in duplicate, following the procedures described by the Association of Official Analytical Chemists (**AOAC**): TN (AOAC, 2000; method 991.20), NPN (AOAC, 2000; method 991.21), and NCN (AOAC, 2000; method 998.05). TP was calculated by subtracting NPN from TN and multiplying the result by the factor 6.38. CN was calculated by subtracting NCN from TN and multiplying the result by the factor 6.38; and then CN%TP was calculated. All milks were also subject to a screening analysis using mid-IR milk spectrophotometric determination of milk components (Lactoscope FTA, Delta Instruments, Drachten, The Netherlands) and a flow cytometric somatic cell count (Somasmart, Delta Instruments).

SDS-PAGE Analysis. The polyacrylamide gels used to determine the trace quantity (mm x OD) of all protein bands (i.e., intact CN, CNPP, and SP), were made following the procedure described by Verdi et al. (1987) using the gel unit Protean^{TM} II

and the power unit 3000/300 from Bio-Rad (Bio-Rad Laboratories Inc., Hercules, CA) gel electrophoresis system. The preparation of the samples for SDS-PAGE was done by diluting each milk with 0.9 g of a buffer that consisted of 10 mM Tris-HCl, 1.0% SDS, 20% glycerol, 0.02% bromophenol blue tracking dye, and 45 mM dithiothreitol. The weight of milk used for the dilution was calculated based on Kjeldahl TP content of the milk sample to achieve a loading of approximately 28 μg of true protein loaded per slot. Each milk plus buffer mixture was placed in a sealed glass vial (Target DP Vials C4000-1W, National Scientific Co., Rockwood, TN), heated to 100°C using steam, held at 100°C for 3 min and stored frozen at -80°C. On the day of the SDS-PAGE run, diluted milks were thawed using steam heated to 100°C for 3 min, and then cooled to about 25°C. Each milk plus buffer mixture was loaded into a different well of the gel. The volume loaded was chosen to achieve a maximum peak OD in the range of 1.0 to 1.4 for the most intense protein band [i.e., α_s-caseins (**α_s-CN**)] in the sample to avoid overloading and nonlinear response in quantitation. One raw milk was used as a marker in one lane of each gel for evaluation of loading consistency and comparison of resolution of the proteins from gel-to-gel. Three SDS-PAGE runs were performed, each with 2 gels. All gels were scanned with USB GS 800 Densitometer (Bio-Rad Laboratories Inc., Hercules, CA) and quantified using Quantity 1 1-D Analysis software.

For quantitative analysis, a center line for each lane on each gel was created during the scanning of the gel. The lines were adjusted individually to match the middle section of the bands in each lane. Next, each lane background was adjusted using the rolling disk method of subtraction to obtain a straight baseline for each lane. The

detection of bands was done by setting the bracket width of the bands and then by adjusting each band in each lane by moving the brackets up or down based on visual observation. The trace quantity (mm X OD) of the intact CN, CNPP, and SP bands, and the sum of all protein bands (i.e., CN+CNPP+SP), of the gel for each milk was calculated. The estimation of CN%TP by SDS-PAGE was calculated in two ways: 1) using only the intact CN bands divided by the total of all protein bands (i.e., CN%TP) and 2) including the sum of intact CN and CNPP bands divided by the total of all protein bands (i.e., CN+CNPP%TP).

Statistical Analysis. Of the 64 milks, 1 was removed from the data set. Criteria for removal of outlier was based on the very high somatic cell count (i.e., 3,330 x 10^3 cells/mL) of one cow for a specific week. A t-test was performed using JMP (JMP, Version Pro 13, SAS Institute Inc., Cary, NC) to determine if there was difference ($P < 0.05$) between the mean CN%TP determined by Kjeldahl and the two estimations given by SDS-PAGE: 1) CN%TP, and 2) CN+CNPP%TP.

Quantification of CN, CNPP, and SP (%) in Kjeldahl NCN filtrate and NCN precipitate

Selection of Milks. Three milks containing different fat, lactose, TP, CN, and SP content were formulated (Kaylegian et al., 2006) as part of the normal production of the modified milk calibration samples. Because the ingredients and approach used to formulate these 3 modified milks (Kaylegian et al., 2006) it was expected that the protein content of the 3 milks would be different, but the CN%TP determined by Kjeldahl analysis would be virtually the same. The 3 milks were analyzed in duplicate

for total solids, lactose, and fat using forced-air oven drying (AOAC International, 2000; method 990.20; 33.2.44), enzymatic lactose (AOAC International, 2000; method 984.15; 33.2.67, Lynch et al., 2007), and ether extraction (AOAC International, 2000; method 989.05; 33.2.26), respectively.

Kjeldahl Analysis. Kjeldahl TN, NPN and NCN methods were performed for each formulated milk, in duplicate, following the procedures described by the AOAC, as indicated above. For the Kjeldahl NCN analysis, the volume of acetic acid, sodium acetate, and water were adjusted according to the TP level, as recommended by Wojciechowski and Barbano (2015). The 6 Kjeldahl NCN filtrates (i.e., 3 milks in duplicate) and 6 NCN precipitates were collected to be analyzed by SDS-PAGE.

SDS-PAGE Analysis. Each of the 6 filtrates and the 6 precipitates resulting from Kjeldahl NCN analysis of the 3 milks, in duplicate, was collected and prepared for SDS-PAGE analysis, in duplicate. The preparation of each NCN filtrate for SDS-PAGE analysis started with the pH adjustment from 4.6 to 6.8, using 1N NaOH or 0.1N NaOH. Next, each filtrate was diluted in a buffer that consisted of 10 mM Tris-HCl, 1.0% SDS, 20% glycerol, 0.02% bromophenol blue tracking dye, and 45 mM dithiothreitol. The weight of filtrate and buffer used for dilution was calculated based on TP content of each filtrate using the results given by Kjeldahl TN and NPN, assuming that all proteins in the NCN filtrate were SP. The target was to have filtrate + buffer mixtures with TP content of 0.3μg per μL. Each NCN filtrate + buffer mixture was placed in a sealed glass vial, heated to 100°C using steam, held at 100°C for 3 min and stored frozen at -80°C.

For the preparation of each NCN precipitate for SDS-PAGE analysis, the precipitate was diluted into a buffer that consisted of 10 mM Tris-HCl, 1.0% SDS, 20%

glycerol, 0.02% bromophenol blue tracking dye, and 45 mM dithiothreitol. The weight of NCN precipitate used for dilution was calculated based on TP content of each precipitate using the information given by Kjeldahl TN and NPN analysis of formulated milk, assuming that all proteins in the NCN precipitate are CN and CNPP. The target was to have precipitate + buffer mixtures with TP content of 8μg per μL. Each NCN precipitate + buffer mixture was placed in a sealed glass vial, heated to 100°C using steam, held at 100°C for 3 min and stored frozen at -80°C.

On the day of the SDS-PAGE run, diluted NCN filtrate and NCN precipitates were thawed using steam heated to 100°C for 3 min, and then cooled to about 25°C. Each sample was loaded into a well of the gel. For the NCN filtrate + buffer samples, the volume loaded was chosen to achieve a max OD in the range of 0.3 to 0.4 for the most intense protein band (i.e., beta-lactoglobulin **(β-LG)**). For the NCN precipitate + buffer samples, the volume loaded was chosen to achieve a max OD in the range of 0.7 to 0.9 for the most intense protein band (i.e., α_s-CN). It is assumed that the amount of Coomassie blue dying bond per unit of protein is the same for CN and SP. To verify this assumption, a SDS-PAGE analysis was performed using different concentrations of MCC and SPI (5, 10, 15, 20. 25, and 30 μg of protein loaded per well) (Nelson and Barbano, 2005). The dye binding rate (dye bound per unit weight of protein interacted with SDS) of Coomassie blue to CN was not significantly different than the binding rate to SP under the conditions used in SDS-PAGE analysis (data not shown). Raw milk was used as a marker on each gel for evaluation of loading consistency and resolution of the proteins from gel-to-gel. Two gels (i.e., front and back) were run and scanned with USB GS 800 Densitometer and quantified using Quantity 1 1-D Analysis software as

described above. The relative quantity (%) of milk immunoglobulin **(IGG)**, bovine serum albumin **(BSA)**, lactoferrin **(LFR)**, α_s -CN, beta-casein **(β-CN)**, kappa-casein **(κ-CN)**, β-LG, alpha-lactalbumin **(α-LA)**, and CNPP for each NCN filtrate buffer mixture and each NCN precipitate buffer mixture was estimated. The relative quantities (%) given by SDS-PAGE were used to calculate the proportion of CN, CNPP, and SP from milk that goes into the NCN filtrate and the proportion of CN, CNPP, and SP that stays in the NCN precipitate.

Statistical Analysis. A t-test was performed using JMP (JMP, Version Pro 13, SAS Institute Inc., Cary, NC) to determine if the proportion of CN, CNPP, and SP (%) from milk that goes into the Kjeldahl NCN filtrate was different ($P < 0.05$) than the proportion of CN, CNPP, and SP (%) from milk that stays in the Kjeldahl NCN precipitate.

RESULTS AND DISCUSSION

Determination of CN%TP by Kjeldahl versus CN%TP by SDS-PAGE

A typical SDS-PAGE gel separation of milk proteins is shown in Figure 3.1 for milk from two cows included in our study. The milk from cow number 2 has much more proteolytic damage to CN with lower (67.02%) intact casein (i.e., α-CN + β-CN + κ-CN) than for cow number 1 (76.34%). The casein proteolysis product identified as CNPP6 makes a large contribution to the proteolysis products in the milk from cow #2 (Figure 3.1). The comparison of mean and standard deviation (SD) of CN%TP estimated by Kjeldahl and CN%TP estimated by SDS-PAGE is shown in Table 3.1. The mean CN%TP estimated by Kjeldahl was higher ($P < 0.05$) than the mean CN%TP estimated

by SDS-PAGE calculated using only intact CN divided by the total of all protein bands (Table 3.1). However, no difference was detected ($P > 0.05$) between Kjeldahl CN%TP (mean = 81.05%) and SDS-PAGE CN%TP (mean = 80.42%) calculated using the sum of intact CN and CNPP divided by the total of all protein bands (Table 3.1). The correlation of Kjeldahl CN%TP with SDS-PAGE CN%TP is shown in Figure 3.2. About 79% of the variability in mean Kjeldahl CN%TP was explained by its relationship with SDS-PAGE CN+CNPP%TP.

The fact that the Kjeldahl estimated CN%TP agreed better with SDS-PAGE measurement of CN%TP when the CNPP were included in the calculation indicates that a substantial proportion of the CNPP may be precipitating with the intact casein in the Kjeldahl NCN method (Table 3.1). However, when there is proteolysis of casein it is not clear how much of the CNPP remain with the NCN precipitate and how much is soluble in the NCN filtrate. Therefore, the second part of the present study was conducted to estimate the proportion of casein proteolysis products that precipitate with intact casein in the Kjeldahl NCN method.

Figure 3.1. SDS-PAGE analysis of milk from each of 2 mid-lactation Holstein cows collected twice in each of two weeks: (A) Cow 1: milk with low casein proteolysis products (CNPP); and (B) Cow 2: milk with high CNPP. Mean relative quantity (%) of lactoferrin (LFR), bovine serum albumin (BSA), immunoglobulin G (IGG), alpha-casein (α-CN), beta-casein (β-CN), CNPP 2, kappa-casein (κ-CN), CNPP 3, beta lactoglobulin (β-LG), alpha-lactalbumin (α-LA), CNPP 6, sum of intact CN (i.e., α-CN + β-CN + κ-CN) , sum of CNPP (i.e., CNPP 2 + CNPP 3 + CNPP 6), and CN+CNPP given by SDS-PAGE.

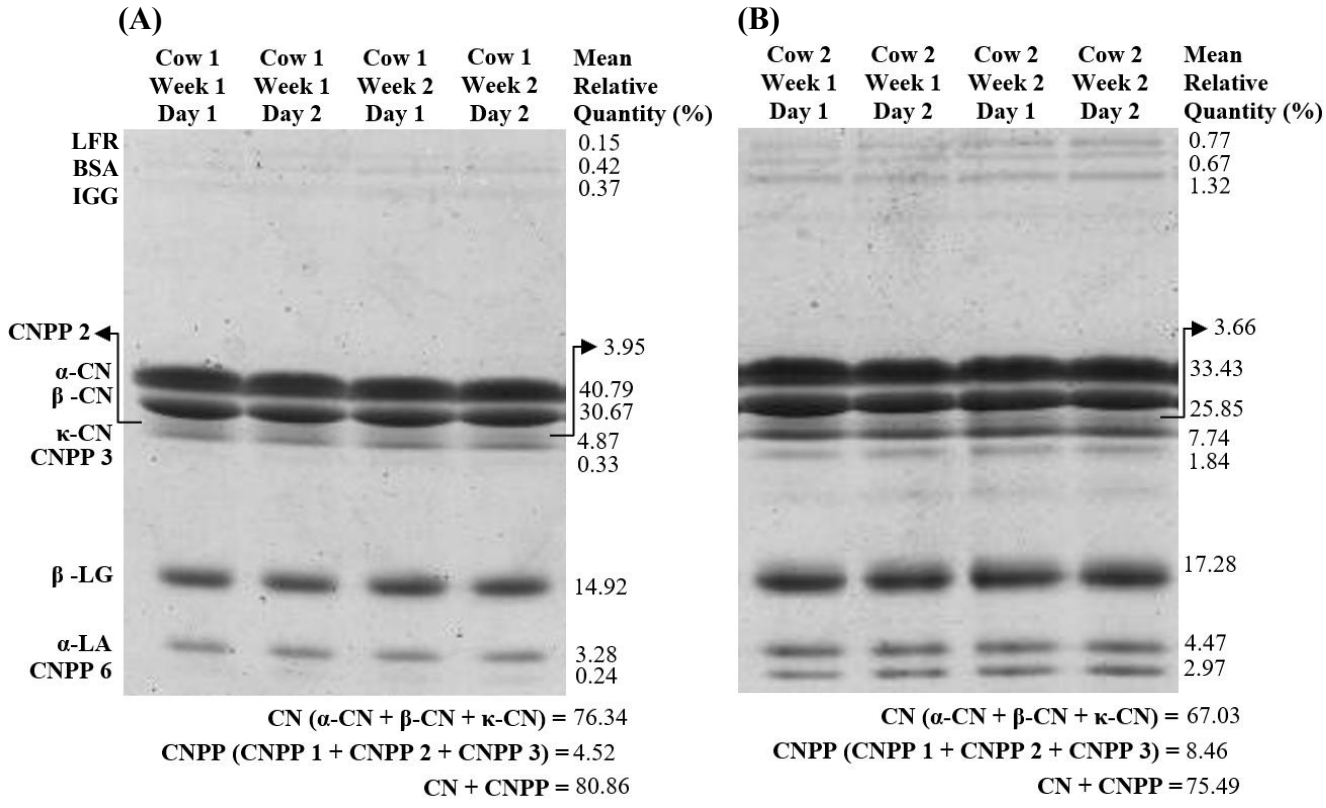
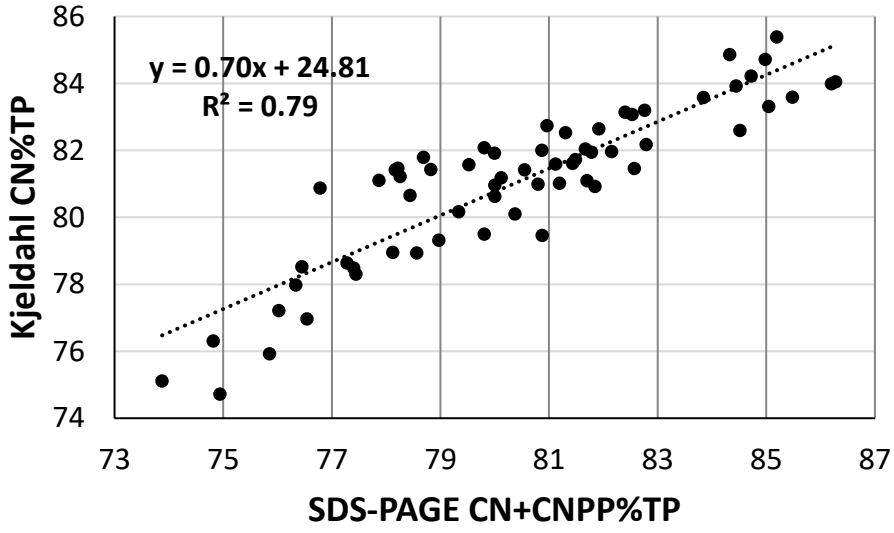


Table 3.1. Mean casein as percentage of true protein (CN%TP), and standard deviation (SD) of milk samples from 16 Holstein cows collected twice every two weeks for a total of 63 milks, estimated by Kjeldahl. SDS-PAGE estimated CN%TP was calculated in two ways: 1) using only intact CN divided by the total of all protein bands and 2) using the sum of CN and casein proteolysis products (CNPP) divided by the total of all protein bands.

	Mean CN%TP	SD
Kjeldahl	81.05[a]	2.35
SDS-PAGE (intact CN only)	75.80[b]	4.37
SDS-PAGE (CN + CNPP)	80.42[a]	2.99

[ab] Means not sharing a common superscript differ ($P < 0.05$)

Figure 3.2. Kjeldahl casein as percentage of true protein (CN%TP) plotted as a function of SDS-PAGE CN%TP calculated using the sum of mm x OD (optical density) intact caseins (CN) plus casein proteolysis products (CNPP) divided by the sum of mm x OD for all protein bands detected, for 63 milks collected from 16 mid-lactation Holstein cows in two different weeks.



Quantification of CN, CNPP, and SP (%) in Kjeldahl NCN Filtrate and NCN Precipitate

The chemical composition of the three formulated milks used for Kjeldahl NCN to produce the NCN filtrates and NCN precipitates is shown in Table 3.2. The composition of the three formulated milks were different, with varying levels of fat (0.19, 2.76, and 5.76%), lactose (4.02, 4.57, and 5.11%), and TP (4.26, 3.24, and 2.05). However, the three milks had about the same CN%TP (81.18, 81.10, 81.57).

We found using SDS-PAGE analysis of the Kjeldahl NCN filtrates and NCN precipitates that out of all intact milk CN, κ-CN was the only one found in both NCN filtrate and precipitate (Table 3.3). Wojciechowski and Barbano (2015) reported the presence of κ-CN in NCN filtrates with pH of 4.48 and 4.79. Furthermore, there were only two SP present in the precipitate: β-LG and BSA. Because our formulated milks

were pasteurized, some thermal denaturation of β-LG occurred due to rupture of intramolecular disulfide bonds and subsequent formation of intermolecular disulfide bonds between β-LG and κ-CN (Zittle et al., 1962; Aurand et al., 1963). Hartman Jr. and Swanson (1965) also found that a stable complex was formed when β-LG was heated with κ-CN. However, a stable complex was not formed when BSA was heated with κ-CN. The author suggests that heated BSA precipitates with CN at pH 4.7 and does not interact with CN during heating. The IGG and LFR were present in the filtrate but not in the precipitate (Table 3.3) and this might be because the isoelectric point of these proteins are higher than 4.6, which is the isoelectric point of caseins (Josephson et al., 1972; Liang et al., 2011).

To facilitate the interpretation of our SDS-PAGE results in comparison to Kjeldahl data, the proteins were divided in three main groups: intact CN (i.e., α-CN + β-CN + κ-CN), CNPP, and SP (i.e., IGG + BSA + LFR + β-LG + α-LA), expressed as mean relative quantity (%), shown in Table 3.4. Using the SDS-PAGE results from Table 3.4, we estimated the mean relative quantity (%) of the three main proteins in milk used for the Kjeldahl NCN analysis. We found (Table 3.5) that 69.74% of the milk TP was intact CN, 11.41% of milk TP was CNPP, and 18.85% of milk TP was SP.

Table 3.2. Chemical composition (% by weight) of the three formulated milks used for quantification of casein (CN), casein proteolysis products (CNPP), and serum proteins (SP) in Kjeldahl NCN filtrate and NCN precipitate.

	Composition [1]										
Milks	TS	Lactose	Fat	TN	NPN	NCN	TP	CN	SP	CN%TP	SP%TP
1	9.563	4.021	0.192	4.419	0.165	0.965	4.255	3.454	0.801	81.175	18.825
2	11.691	4.573	2.764	3.428	0.185	0.798	3.243	2.630	0.613	81.098	18.902
3	13.963	5.113	5.759	2.228	0.182	0.559	2.046	1.669	0.377	81.574	18.426

[1] TS = total solids; TN = total nitrogen X 6.38; NPN = non-protein nitrogen X 6.38; NCN = non-casein nitrogen X 6.38; TP = true protein (TN – NPN); CN = casein (TN – NCN); SP = serum protein (TP – CN).

Table 3.3. Mean relative quantity (%) of milk immunoglobulin (IGG), bovine serum albumin (BSA), lactoferrin (LFR), alpha-casein (α-CN), beta-casein (β-CN), kappa-casein (κ-CN), beta lactoglobulin (β-LG), alpha-lactalbumin (α-LA), and casein proteolysis products (CNPP), in Kjeldahl non-casein nitrogen (NCN) filtrate and precipitate determined with SDS-PAGE.

Kjeldahl NCN	IGG	BSA	LFR	α-CN	β-CN	κ-CN	β-LG	α-LA	CNPP	Total
Filtrate 1	1.27	ND	2.40	ND	ND	4.21	65.67	19.00	7.45	100
Filtrate 2	1.11	ND	2.44	ND	ND	3.08	67.73	19.13	6.52	100
Filtrate 3	1.59	ND	2.80	ND	ND	2.94	66.79	19.28	6.60	100
Mean	1.32	ND	2.55	ND	ND	3.41[b]	66.73[a]	19.14	6.86[b]	100
SD	0.29	ND	0.43	ND	ND	0.92	1.32	1.07	1.13	0
Precipitate 1	ND	0.92	ND	43.57	32.88	8.43	1.14	ND	13.07	100
Precipitate 2	ND	1.26	ND	43.89	32.57	8.57	1.19	ND	12.53	100
Precipitate 3	ND	1.69	ND	42.29	34.20	8.65	1.39	ND	11.79	100
Mean	ND	1.29	ND	43.25	33.22	8.55[a]	1.24[b]	ND	12.46[a]	100
SD	ND	0.39	ND	1.74	1.91	0.85	0.37	ND	1.76	0

ND = not detected

[ab] Means within a column not sharing a common superscript differ ($P < 0.05$)

Table 3.4. Mean relative quantity (%) of intact caseins (CN), casein proteolysis products (CNPP), and serum proteins (SP) in the Kjeldahl non-casein nitrogen (NCN) filtrate and in the precipitate determined with SDS-PAGE.

Kjeldahl NCN	Intact CN	CNPP	SP	CN+CNPP+SP
Filtrate 1	4.21	7.45	88.34	100
Filtrate 2	3.08	6.52	90.40	100
Filtrate 3	2.94	6.60	90.47	100
Mean	3.41[a]	6.86[b]	89.74[b]	100
SD	0.92	1.13	1.62	0
Precipitate 1	84.87	13.07	2.06	100
Precipitate 2	85.03	12.53	2.45	100
Precipitate 3	85.14	11.79	3.08	100
Mean	85.01[b]	12.46[a]	2.53[a]	100
SD	1.77	1.76	0.54	0

[ab] Means in the same column not sharing a common superscript differ ($P < 0.05$)

Table 3.5. Mean and standard deviation (SD) of SDS-PAGE estimated intact casein (CN), casein proteolysis products (CNPP), serum protein (SP), and sum of intact CN, CNPP, and SP (intact CN+CNPP+SP) in formulated milks, expressed as % of true protein (TP) in milk; and mean and SD of Kjeldahl TP nitrogen (%) in 3 formulated milks.

	SDS-PAGE (% of TP in milk)				Kjeldahl
Kjeldahl NCN	Intact CN[1]	CNPP[2]	SP[3]	CN+CNPP +SP	(% of TP nitrogen)
Milk 1	69.69	12.01	18.30	100	100
Milk 2	69.54	11.39	19.07	100	100
Milk 3	69.99	10.83	19.18	100	100
Mean	69.74	11.41	18.85	100	100
SD	1.54	1.41	0.57	0	0

[1]intact CN = intact CN in the NCN filtrate + intact CN in the NCN precipitate

[2]CNPP = CNPP in the NCN filtrate + CNPP in the NCN precipitate

[3]SP = SP in the NCN filtrate + SP in the NCN precipitate

We used the Kjeldahl TN, NPN, and NCN results in combination with the SDS-PAGE results to calculate the % of intact CN, CNPP, and SP from milk that goes into the Kjeldahl NCN filtrate and Kjeldahl NCN precipitate (Table 3.6). About 81.28% of milk TP was CN, based on Kjeldahl results. However, we found using SDS-PAGE that out of 81.28% only 69.10% was intact CN, and 10.13% was CNPP (Table 3.6). It can be seen from the scan of the SDS-PAGE gel (Figure 3.3) that the CNPP6 band made the largest contribution in the CNPP % value estimated for the NCN precipitate. Trieu-Cuot and Gripon (1981) reported that this band consists of γ_2-CN and γ_3-CN, and because of their high hydrophobicity, these two peptides precipitate with CN in the Kjeldahl NCN sample preparation method. Based on SDS-PAGE results, 11.41% of the milk TP was CNPP (Table 3.5), and a majority of these CNPP (10.13% of milk TP) went into the NCN precipitate, while only a small fraction (1.28% of milk TP) remained in the NCN filtrate (Table 3.6). Therefore, the portion of CNPP from formulated milk that stays with the NCN precipitate during Kjeldahl NCN analysis is significantly different ($P < 0.05$) than the CNPP that goes into the filtrate. The Kjeldahl CN%TP metric for a milk sample underestimates the amount of enzymatic damage that has happened to casein because most of the casein proteolysis products of casein are retained in the NCN precipitate and counted as casein. The functionality, product yield, and flavor implications of this are different for different dairy products.

While a small amount of intact CN remained in the NCN filtrate, a relatively larger amount of CNPP and SP goes into the NCN precipitate, increasing the Kjeldahl CN%TP value (Table 3.6). The net effect is an overestimation of CN%TP by Kjeldahl when CN is defined as only intact casein in SDS-PAGE without the inclusion of CNPP.

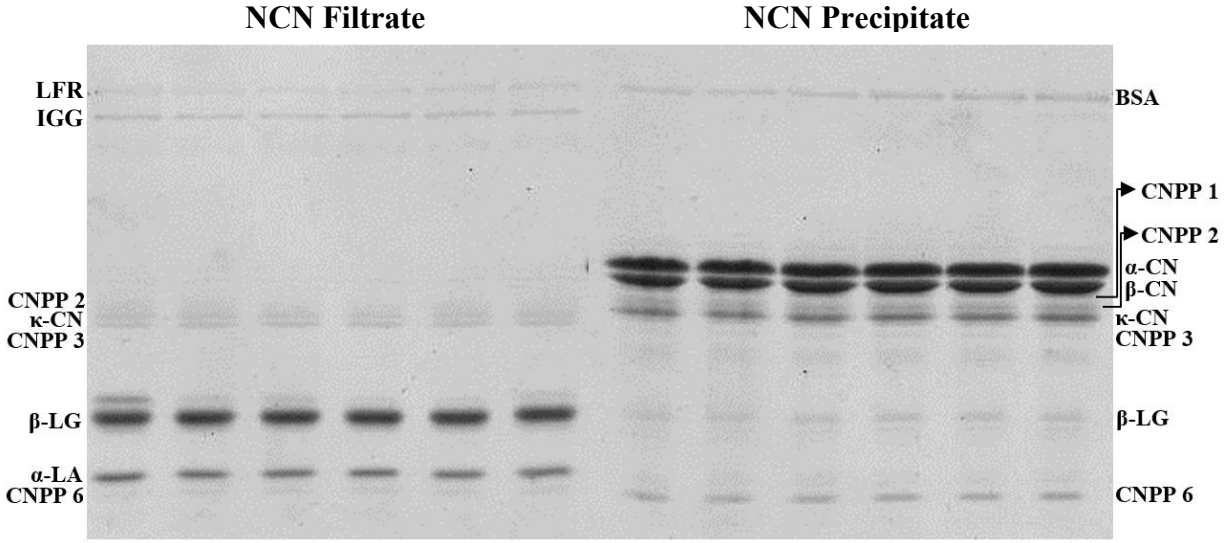
A rapid method to measure CNPP in milk is needed to reflect the amount of proteolysis of casein and the quality of milk protein and this may be achieved by development of a partial least squares (**PLS**) model to quantify CNPP using a MIR.

Table 3.6. Mean and standard deviation (SD) of SDS-PAGE estimated casein (CN), casein proteolysis products (CNPP), serum protein (SP), and sum of CN, CNPP, and SP (CN+CNPP+SP) in the Kjeldahl non-casein nitrogen (NCN) filtrates and NCN precipitates, expressed as % of true protein (TP) in milk; and mean and SD of Kjeldahl TP nitrogen (%) in the Kjeldahl NCN filtrates and NCN precipitates.

	SDS-PAGE (% of TP in milk)				Kjeldahl
Kjeldahl NCN	Intact CN	CNPP	SP	CN+CNPP+SP	(% of TP nitrogen)
Filtrate 1	0.79	1.40	16.63	18.83	18.83
Filtrate 2	0.58	1.23	17.09	18.90	18.90
Filtrate 3	0.54	1.22	16.67	18.43	18.43
Mean	0.64[b]	1.28[b]	16.80[a]	18.72[b]	18.72
SD	0.18	0.21	0.32	0.22	0.22
Precipitate 1	68.90	10.61	1.67	81.18	81.18
Precipitate 2	68.96	10.16	1.99	81.10	81.10
Precipitate 3	69.45	9.62	2.51	81.57	81.57
Mean	69.10[a]	10.13[a]	2.06[b]	81.28[a]	81.28
SD	1.46	1.43	0.45	0.22	0.22

[ab] Means in the same column not sharing a common superscript differ ($P < 0.05$)

Figure 3.3. SDS-PAGE gel loaded using 3 Kjeldahl NCN filtrate and 3 Kjeldahl NCN precipitate, in duplicate. From left to right: filtrate 1 (R1), filtrate 1 (R2), filtrate 2 (R1), filtrate 2 (R2), filtrate 3 (R1), filtrate 3 (R2), precipitate 1 (R1), precipitate 1 (R2), precipitate 2 (R1), precipitate 2 (R2), precipitate 3 (R1), and precipitate 3 (R2), where R = replicate. LFR = lactoferrin; BSA = bovine serum albumin; IGG = immunoglobulin G; α-CN = alpha-casein; β-CN = beta-casein; CNPP = casein proteolysis products; κ-CN = kappa-casein; β-LG = beta lactoglobulin; α-LA = alpha-lactalbumin.



CONCLUSIONS

The purpose our study was to compare the Kjeldahl estimate of CN%TP with the SDS-PAGE estimate of CN%TP for the same milk and to determine the proportion of CN, CNPP, and SP from milk TP that goes into each of the Kjeldahl NCN filtrate and precipitate. We found that the estimate of CN%TP by Kjeldahl was significantly higher ($P < 0.05$) than the estimate of CN%TP by SDS-PAGE that was calculated as only intact CN divided by the total of all protein bands. However, no significant difference ($P > 0.05$) in the estimate of CN%TP by Kjeldahl was detected when compared to CN%TP by SDS-PAGE when CNPP were included as CN in the calculation of SDS-PAGE results. Thus, a considerable proportion of CNPP were

retained in the NCN precipitate. The Kjeldahl NCN method assumes that only CN is precipitated while the remaining NCN portion of milk (including SP) remained in the filtrate. Based on SDS-PAGE results, we found that a majority (89%) of the CNPP from the milk (approximately 10.13% out of 11.41% TP) were retained in the Kjeldahl NCN precipitate, and the largest contribution to CNPP in the precipitate was CNPP6. Thus, the classical Kjeldahl NCN analysis underestimates the amount of proteolytic enzyme damage that has happened to casein.

ACKNOWLEDGEMENTS

The authors thank the Miner Institute (Chazy, NY) for their collaboration by providing individual cow milks for use in the study. The authors thank the Northeast Dairy Foods Research Center (Ithaca, NY) for support of this research. The technical assistance of Chassidy Coon, Michelle Billota, Larissa Di Marzo, and Sara Bova of Cornell University was greatly appreciated.

REFERENCES

AOAC. 2000. Official Methods of Analysis. 17th ed. AOAC, Gaithersburg, MD.

Aurand, L. W., J W. Brown, and J. G. Lecce. 1963. Effect of Heat on the Proteins of Milk as Revealed by Gel and Immunoelectrophoresis. J. Dairy Sci. 46:1177-1182.

Barbano, D. M., J. L. Clark, C. E. Dunham, and J. R. Fleming. 1990. Kjeldahl method for determination of total nitrogen content of milk: collaborative study. J. AOAC.73:849-859.

Barbano, D. M., J. M. Lynch, and J. R. Fleming. 1991. Direct and indirect determination of true protein content of milk by Kjeldahl analysis: Collaborative study. J. AOAC. 74:281-288.

Barbano, D. M., and J. M. Lynch. 1992. Crude and protein nitrogen bases for protein measurement and their impact on current testing accuracy. J. Dairy Sci. 75:3210-3217.

Fox, P. F., T. Uniacke-Lowe, P. L. H. McSweeney, and J. A. O'Mahony. 2015. Biologically Active Compounds in Milk: Bioactive Milk Proteins. Page 424 in Dairy Chemistry and Biochemistry. Springer, Switzerland.

Hartman Jr., G. H., and A. M. Swanon. 1965. Changes in mixtures of whey protein and k-casein due to heat treatments. J. Dairy Sci. 48:1161-1167.

Josephson, R. V., E. M. Mikolajick, D. P., Sinha. 1972. Gel isoelectric focusing of selected bovine immunoglobulins. J. Dairy Sci. 55:1508-1510.

Kaylegian, K. E., G. E. Houghton, J. M. Lynch, J. R. Fleming, and D. M. Barbano. 2006. Calibration of infrared milk analyzers: modified milk versus producer milk. J. Dairy Sci. 89:2817–2832.

Karman, A. H., and M. A. J. S., van Boekel. 1986. Evaluation of the Kjeldahl factor for conversion of the nitrogen content of milk and milk products to protein content. Neth. Milk Dairy J. 40:315-336.

Liang Y., X. Wang, M. Wu, and W. Zhu. 2011. Simultaneous isolation of lactoferrin and lactoperoxidase from bovine colostrum by SPEC 70 SLS cation exchange resin. Int. J. Environ. Res. Public Health. 8:3764-3776.

Lynch, J. M., and D. M. Barbano. 1999. Kjeldahl Nitrogen Analysis as a Reference Method for Protein Determination in Dairy Products. J. AOAC Int. 82:1389-1398.

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.

Margolies, B., M. C. Adams, J. Pranata, K. Gondoutomo, and D. M. Barbano. 2017. Effect of uncertainty in composition and weight measures in control of cheese yield and fat loss in large cheese factories. J. Dairy Sci. 100:6822-6852.

Nelson, B. K., and D. M Barbano. 2005. A Microfiltration process to maximize removal of serum proteins from skim milk before cheese making. J. Dairy Sci. 88:1891-1900.

Rowland, S. J. 1938a. The precipitation of the proteins in milk: I. Casein II. Total proteins III. Globulin IV. Albumin and proteosepeptone. J. Dairy Res. 9:30-41.

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

Trieu-Cuot, P., and J. Gripon. 1981. Electrofocusing and two-dimensional electrophoresis of bovine caseins. J. Dairy Res. 48:303-310.

USDA. 2017. Announcement of advanced prices and pricing factors. Accessed Apr. 27, 2017. https://www.ams.usda.gov/mnreports/dymadvancedprices.pdf.

Verdi, R. J., D. M. Barbano, and M. E. Dellavalle. 1987. Variability in true protein, casein, nonprotein nitrogen, and proteolysis in high and low somatic cell count milks. J. Dairy Sci. 70:230–242.

Wojciechowski, K. L. and D. M. Barbano. 2015. Modification of the Kjeldahl noncasein nitrogen method for bovine milk concentrates and milks from other species. J. Dairy Sci. 98:7510-7526.

Wrolstad, R. E., T. E. Acree, E. A. Decker, M. H. Penner, D. S. Reid, S. J. Schwartz, C. F. Shoemaker, D. Smith, P. Sporns. 2005. Characterization of proteins. Pages 155-265 in Handbook of Food Analytical Chemistry. John Wiley & Sons, Inc., Hoboken, New Jersey.

Zittle, C. A., M. P. Thompson, J. H. Custer, and J. Cerbulis. 1962. κ-casein – β-LG interaction in solution when heated. J. Dairy Sci. 45:807-810.

CHAPTER FOUR

CONCLUSIONS AND FUTURE WORK

Mid infrared (**MIR**) milk analyzers are secondary testing instruments that are calibrated by reference chemistry methods for rapid analysis of milk components. Currently, MIR milk analyzers are calibrated using the Kjeldahl method for the routine determination of protein concentration. MIR instruments measure the infrared light absorbance of the components' molecular vibrations, and a calibration model correlates this to the components' concentrations. Different milk proteins are expected to exhibit unique infrared light absorbance characteristics due to differences in their structure and amino acid composition, so it should be possible to develop a MIR calibration model to predict the concentration of fractions of milk proteins, which includes casein (**CN**) and serum protein (**SP**).

Partial least squares (**PLS**) models using MIR spectra to predict true protein (**TP**), CN, SP, and CN as a percentage of TP (**CN%TP**) content of micellar casein concentrates (**MCC**) and unflavored milk-based beverages were developed and validated. The PLS models, developed using 625 samples of formulated milk with a range of CN%TP and orthogonal matrix of fat, TP, and lactose, yielded ratio of performance to deviation (**RPD**) values of 62.7, 31.4, 14.1, and 3.8 for the prediction of TP, CN, SP, and CN%TP, respectively. External validation of the PLS models using 18 formulated MCC with three different levels of lactose and varying CN%TP resulted in relative standard deviation (**RSD**) of 0.558, 0.898, 4.257, and 0.591% for prediction of TP, CN, SP, and CN%TP, respectively. The RSD for prediction of

CN%TP in the MCC validation sample set using a calculated value from TP and CN was lower (0.537 vs 0.591%) than the direct PLS model to predict CN%TP. External validation test of the PLS model using 28 formulated unflavored milk-based beverage with three different levels of fat and varying CN%TP resulted in RSD of 0.350, 1.637, 4.049, and 10.887% for prediction of TP, CN, SP, and CN%TP, respectively. The RSD for prediction of CN%TP in the unflavored milk-based beverage validation sample set using a calculated value from TP and CN was much lower (1.958 vs 10.89%) than the direct PLS model to predict CN %TP. Some, or all, of the concentrations of TP, CN, and SP in the populations of MCC and milk-based beverage validation samples were outside the range of concentrations of those milk components in the population of samples used to develop the PLS models, yet the validation performance of the PLS models was very good. The ability of these PLS models to perform well on these sample matrices may be due to the high RPD values achieved by orthogonal design of the modeling population used for the PLS model development.

The estimate of CN%TP measured using Kjeldahl was compared to the estimate of CN%TP measured using sodium dodecyl sulfate polyacrylamide gel electrophoresis (**SDS-PAGE**) for the same milk, and the proportion of CN, casein proteolysis products (**CNPP**), and SP from milk TP that goes into each of the Kjeldahl non-casein nitrogen (**NCN**) filtrate and precipitate were determined. The estimate of CN%TP by Kjeldahl was significantly higher ($P < 0.05$) than the estimate of CN%TP by SDS-PAGE that was calculated as only intact CN divided by the total of all protein bands. However, no significant difference ($P > 0.05$) in the estimate of CN%TP by

Kjeldahl was detected when compared to CN%TP by SDS-PAGE when CNPP were included as CN in the calculation of SDS-PAGE results. Thus, a considerable proportion of CNPP were retained in the NCN precipitate. The Kjeldahl NCN method assumes that only CN is precipitated while the remaining NCN portion of milk (including SP) remained in the filtrate. Based on SDS-PAGE results, a majority (89%) of the CNPP from the milk (approximately 10.13% out of 11.41% TP) were retained in the Kjeldahl NCN precipitate, and the largest contribution to CNPP in the precipitate was CNPP6. Thus, the classical Kjeldahl NCN analysis underestimates the amount of proteolytic enzyme damage that has happened to casein.

In cheesemaking, the estimate of CN which includes intact CN and CNPP remaining in the Kjeldahl NCN precipitate may have practical usefulness as the protein content in cheese is determined by the milk protein that precipitates at approximately pH 4.6, similar to the Kjeldahl NCN principles. However, in MCC and milk-based beverages, the definition of CN as only intact CN (not including CNPP) would be more useful because the content of CNPP indicates the extent of proteolytic damage, which is undesirable in these products. Therefore, a rapid method to measure CNPP in milk and milk-based products is needed to reflect the amount of proteolysis of caseins and the quality of milk protein. This may be achieved by development of a PLS model to quantify CNPP using a MIR milk analyzer.

In the future, the TP, CN, SP, and CN%TP MIR models could be used as a rapid method for measurement of protein composition and quality assurance of milk and milk-based beverage products. At production facilities using near-line MIR milk analyzers, this method can provide routine analysis of milk protein composition that is

usually not feasible using the more time consuming and expensive Kjeldahl method. Additionally, a CNPP MIR model should be developed for MIR milk analyzers so that an accurate measurement of proteolytic damage to casein as an indicator of milk protein quality can also be performed rapidly and economically.