Developing an optical backscatter method for determining casein micelle particle size in heated milk

Heather Taterka, Anna Zamora [*], Manuel Castillo [*]

Centre d'Innovació, Recerca i Transferència en Tecnologia dels Aliments (CIRTTA), Departament de Ciència Animal i dels Aliments, Facultat de Veterinària, Universitat Autònoma de Barcelona, 08193 Bellaterra, Cerdanyola del Vallès, Barcelona, Spain

ARTICLE INFO

ABSTRACT

A plethora of different factors, such as heat treatment, pH, soluble calcium and phosphate concentrations, colloidal calcium phosphate, ionic strength, redox potential, etc., affect functionally of critical milk components such as casein micelles, fat globules and whey proteins. These physicochemical changes induce fat- or protein-protein interactions that would be associated to changes in particle size that might be revealed using light backscatter measurements. We hypothesized that inline, simple, low-cost light backscatter measurements might have the potential to provide functionally related information, representing an interesting opportunity for process control. Casein micelle particle size and near infrared light backscatter spectra were measured in milks heat treated at 80 and 90 °C and pH 6.3, 6.7 and 7.1 in order to obtain prediction models for estimating changes in casein micelle particle size during milk heat treatment. Light intensity was measured over a spectral range of 200–1100 nm using a simple optical backscatter sensor and was implemented into models for particle size predictions as a function of heat treatment temperature and pH. Models which included an exponential factor containing a ratio of two specific wavebands were found to improve R^2 when compared to single wavelength models. The best model exhibited an R^2 of 0.993 and SEP of 2.36 nm. The developed prediction models show promise for in-line monitoring of whey protein denaturation and casein micelle particle size.

1. Introduction

In-, on– and at-line monitoring of liquid milk products has implementation potential in the dairy industry for process control and time and cost savings. To date, several applications have been developed to that end, but it is still plenty of room for innovative process analytical technologies that could provide useful, real-time information for process control optimization and improvement. In that sense, the monitoring of various milk product manufacturing steps such as determination of milk gelation mechanisms and changes during the cheese making process has been investigated using light scatter techniques (Castillo et al., 2005c; Castillo et al., 2006; Fagan et al., 2007). Near infrared (NIR) spectroscopy has been widely used at- or even in-line in milk to determine absorption in the infrared region (780–2500 nm) of bonds and chemical groups to quantify various milk components (García, 2004). Robert et al. (1987) was able to establish specific wavelengths corresponding to changes in fat, protein and lactose content in milk, however noted that some interference was observed as a result of large water absorption and light scattering of fat particles. As well, it was observed by Diaz-Carrillo et al. (1993) that NIR spectroscopy could be used to successfully quantify protein, fat and total casein in goat́s milk. NIR spectroscopy has also been used to develop prediction equations for the determination of a number of milk attributes with in-line application potential. For example, for the differentiation of different heat treatments in milk (Downey et al., 1990), detection of adulteration by the addition of whey powder to milk powders (Giangiacomo et al., 1991) or addition of NaCl and skim milk powder to milk (Pedretti et al., 1993). Sørensen and Jepsen (1997) used NIR spectroscopy to detect cheese defects as a result of *Clostridium tyrobutyricum*.

Development of prediction models has also been accomplished using NIR and light scatter techniques to predict cheese making characteristics at real time. Castillo, et al. (2000) used NIR light backscatter for developing simple prediction models using only optical parameters for the determination of cutting and clotting time. Further developments have allowed the in-line transformation of the NIR light backscatter response during milk coagulation into real time and continuous prediction of the gel elastic modulus during cheese manufacturing (Arango and Castillo, 2018) and pH during yogurt fermentation (Arango et al.,

---

2020). Light sidescatter and transmission were also used to estimate whey fat concentration, where sidescatter produced models with a higher R^2 (greater than 0.95) than corresponding transmission models (less than 0.5) (Castillo et al., 2005b). An in-line syneresis sensor was developed by Castillo, et al. (2007), which was applied by Fagan, et al. (2008) to model cheese manufacturing indices such as whey fat, curd yield and curd moisture content using light backscatter sensor technology achieving R^2 values of 0.90 or more for the prediction models. In particular, their work was mainly aimed toward the determination of curd moisture content as a function of time in order to improve final moisture content during cheese making. Initial models were determined from a total of 40 parameters; however successful models utilized a combination of the parameters: temperature, percentage of protein, milk fat, and milk solids as well as milk fat protein ratio and light backscatter intensity ratios, which contain coagulation and syneresis information. Other studies have investigated the relation between light scatter and whey protein denaturation and the subsequent attachment of whey proteins (WPs) to the surface of the casein micelle (Lamb, Payne, Xiong, and Castillo, 2013), however the strong pH effect of this mechanism has not been thoroughly investigated for optical sensor development.

The authors hypothesize that particle size may be used as an indirect measure of whey protein denaturation in heat treated milk. Particularly because the increase of particle size has been shown to be a good indicator of the extent of binding of denatured whey proteins to the surface of the casein micelle (Anema and Li, 2003, Taterka and Castillo, 2015). However, this mechanism is affected to a large degree by milk pH. The maximum attachment level of denatured whey protein to the casein micelle surface has been found to occur at pH $\sim$ 6.3, whereas the formation of soluble whey protein aggregates is the preferential mechanism at higher pH (maximum at pH 7.1) (Vasbinder and de Kruif, 2003; Donato and Guyomarc'h, 2009; Kethireddipalli et al., 2010; Taterka and Castillo, 2015). Moreover, milk pH also affects the light backscatter signal as it has been exhibited that bound/soluble aggregate formation is highly dependent on pH and corresponds significantly to the intensity of light backscatter signal (Taterka and Castillo, 2015). Another important factor is the temperature of heat treatment, which significantly increases the amount of denatured whey protein, resulting in bound and/or soluble serum aggregates, depending on pH. It should be noted that, whey protein denaturation plays a quite significant role in milk and milk products from a functional point of view. For instance, it is well-known that intensive whey protein denaturation impairs milk coagulation during cheese manufacturing while, contrarily, it improves yogurt texture and whey retention. Additionally, intense thermal denaturation is required for thermal stabilization of milk prior to sterilization of unsweetened condensed milk. Since the previous models obtained by Lamb et al. (2013) did not take into account the pH dependence of the binding reaction, the objective of this study was to look further into modeling the light scatter signal with respect to the changes in particle size in heat treated milk, while taking into consideration the effect of milk pH.

2. Materials and methods

2.1. Experimental design

The experiment consisted of a 3 × 2 factorial design with three pH values (6.3, 6.7 and 7.1) and two heat treatments (80 and 90 °C for 10 min). Each treatment was replicated three times.

2.2. Sample preparation

Low-heat skim milk powder (low-heat, spray-dried skim milk powder; pH = 6.5, solubility = 99 %, WPNI $\geq$ 7 mg g^{-1}, 800 cfu g^{-1}) was supplied by Chr. Hansen SL (Barcelona, Spain). Milk was reconstituted with Milli-Q water to a final solids content of 12 % (w/w). Reconstituted milk, initially at pH 6.55 ($\pm$0.015), was pH-adjusted to 6.3, 6.7 and 7.1

at 21 °C using 0.5 M HCl or 0.5 M NaOH. re-adjusted milk samples were allowed to equilibrate for 2 h in a dark place, before final pH reading and minor re-adjustments.

A stainless-steel plate (6 mm slot) containing 80 mL of milk was placed into a water bath (OvanTherm, Suministros Grupo Esper, S.L., Badalona, Spain) controlled using an OvanTherm TC00 unit (resolution 0.1, stability ± 0.1 °C) (Suministros Grupo Esper, S.L.) at either 80 or 90 °C for 10 min. Samples took less than 30 s to reach the target temperature and heat treatment was stopped by removing the heat plate and rapidly placing it in a 0 °C ice-water bath for 3 min. Milk was then transferred to a test tube, refrigerated at 4 °C for no more than 2 days, and re-equilibrated to 21 °C before analysis.

2.3. Casein micelle particle size

Particle size z-average was measured at 20 ± 0.5 °C using dynamic light scattering set to 90° and refractive index set to 1.471 with a Malvern Zetasizer 4 (Malvern Instruments ltd., Malvern, Worcs., UK). Samples were suspended in Ca/imidazole buffer (20 mM-imidazole, 5 mM-CaCl$_2$, 30 mM-NaCl, pH 7.0) at a concentration of 1:2 in order to suspend the casein micelles and allow stability during measurements (Anema, 1997; Anema & Li, 2003; Anema et al., 2004). Measurements were 3 min each and an average of 3 measurements was taken for each representative sampling.

2.4. Light backscatter measurements

Full detail on the optical system used is given in Taterka and Castillo (2015). Briefly, it consisted of a light scatter probe (Reflectronics Inc, Lexington, KY, USA) connected through two 600-μm UV-NIR optical fibers to a tungsten halogen light source (LS-1; Ocean Optics, Inc. Dunedin, FL, USA) and a high-resolution spectrometer (Model HR4000; Ocean Optics, Inc.). Temperature control was achieved by circulating water at 25 °C inside double jacketed walls below the sample well. Spectra Suite software (Ocean Optics, Inc.) was used to collect intensity spectra in the range 200–1100 nm. For single wavelength prediction models, light backscatter maximum intensity (570 nm) was recorded and used for further analysis. In addition, an alternative technique of waveband ratios for modeling proposed by Lamb, Payne, Xiong, and Castillo (2013) was employed (see section 3.2.1 for further detail).

2.5. Statistical analysis

Statistical analysis and interpretation of data was accomplished using "Statistical Analysis System" (SAS, version 9.2, SAS Institute Inc., Cary, NC, USA). Single wavelength models (models 1–5) were developed using the PROC REG function. The development of models 1–5 is more thoroughly described in Section 3.1.

Models 6–10 utilized a technique of waveband ratios using a grouping of two wavebands portions of the spectra, determined by SAS code from Lamb et al. (2013). Wavebands were defined as 35 nm portions of the spectra, thus all grouping of waveband ratios yielded 27 distinct combinations. Prior studies on the behavior of particle size, it has been considered that in a nonlinear regression model the response variable (particle size) is an exponential function as a function of pH (Taterka and Castillo, 2018). The NLIN function was used to form nonlinear regression models, assuming that the response variable follows an exponential model. From this, waveband ratios were developed with the objective of modeling the radius of the casein micelle versus pH and temperature. Models were assessed based on the determination coefficients (R^2) and the standard errors of prediction (SEP) for the regressions between the predicted and actual values. Supplementary information regarding the development of wavelength ratios can be found in Section 3.2.1.

3. Results and discussion

3.1. Single wavelength prediction models

A model for particle size as a function of light backscatter intensity at 570 nm (single wavelength model) was developed. Fig. 1 shows the trends for both light backscatter at 570 nm and z-average particle size as a function of pH and temperature. Models were "separated" between heat treatment temperatures (80 and 90 °C), temperature "integrated" or temperature "combined". Temperature "integrated" models were developed as a function of temperature, whereas temperature combined used all data points but did not include temperature in the prediction equation. Two types of models were obtained: linear models with the form ($y = ax + b$) and quadratic models in the form ($y = ax^2 + bx + c$). As a first modeling approach, both light backscatter and particle size were modeled as a function of pH and/or T as follows:
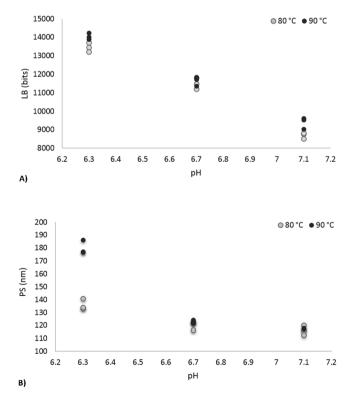
$$LB = \beta_0 + \beta_1 pH \tag{1}$$

$$LB = \beta_0 + \beta_1 pH + \beta_2 T \tag{2}$$

$$PS = \beta_0 + \beta_1 pH + \beta_2 pH^2 \tag{3}$$

$$PS = \beta_0 + \beta_1 pH + \beta_2 pH^2 + \beta_3 T \tag{4}$$

where LB is the light backscatter intensity (bits) measured at the maximum intensity wavelength encountered within the wavelength range between 200 and 1100 nm (i.e., 570 nm), PS is the particle size z-average (nm), T is the heat treatment temperature (°C), and β_{0-3} are regression coefficients.

Light backscatter and particle size experimental data was fit to both linear and quadratic mathematical models to estimate regression coefficients and summary statistics using the PROC REG function in Statistical Analysis System (SAS version 9.2, SAS Institute Inc., Cary, NC, USA). In all cases, higher R^2 values were observed in quadratic models,

however those models which did not show a large difference in R^2 between quadratic and linear models were simplified into the linear model form. Also, in some "integrated" models, temperature (T) was used for the model as the R^2 value increased markedly in some models when including the T variable. Models in which the addition of T did not increase the R^2 value to a large extent were simplified into temperature combined models in order to maintain degrees of freedom (DF) for the model. Table 1 shows the various models that were developed and their summary statistics. Temperature "separated" models and temperature "integrated" models were developed to predict light backscatter maximum intensity as a function of pH and temperature (model 1 and 2) (Fig. 2) and particle size z-average as a function of pH and temperature (model 3 and 4) (Table 1) (Fig. 3). Model 1 utilizes a simple linear model using only pH as a dependent variable, yet still results in high R^2 values for both 80 and 90 °C models ($R^2 = 0.979$ and 0.988, respectively) (Table 1) (Fig. 2A). In the case of the temperature "integrated" model (model 2) (Fig. 2B), all temperature data points were considered (DF = 15) and a high R^2 value of 0.984 was obtained. For the particle size temperature "separated" models (model 3) (Fig. 3A) R^2 values of 0.881 and 0.992 were obtained for 80 and 90 °C models, respectively, whereas the temperature "integrated" model (model 4) (Fig. 3B) had a slightly lower R^2 value of 0.791. In general, it can be seen that light backscatter showed a consistent linearly decreasing trend with pH (Fig. 1A), whereas changes in particle size tended to show a more quadratic response as a function of pH, where 90 °C samples showed greater changes with pH than 80 °C samples (Fig. 1B).

Trends for changes in particle size and light backscatter maximum intensity were investigated and prediction models were developed with high R^2 values as a function of milk pH (Models 1–4). As the technique to measure particle size can be time consuming and costly, the light backscatter method could be a good alternative to particle size

Fig. 1. (A) Light backscatter maximum intensity (LB) as a function of pH at 80 and 90 °C (N = 18); (B) Particle size z-average as a function of pH at 80 and 90 °C (N = 18).

Table 1
Predictive models 1 to 4 for light-backscatter intensity (LB) and particle size z-average (PS).

Model	Prediction equation	Temperature	DF err	Regression coef.	R^2	SEP
1	$LB = \beta_0 + \beta_1 pH$	80 °C	7	$\beta_0 = -5.11 \bullet 10^4$	0.979	320
				$\beta_1 = 5.95 \bullet 10^3$		
		90 °C	7	$\beta_0 = -5.07 \bullet 10^4$	0.988	235
				$\beta_1 = 5.82 \bullet 10^3$		
2	$LB = \beta_0 + \beta_1 pH + \beta_2 T$	Integrated	15	$\beta_0 = 4.68 \bullet 10^4$	0.984	273
		(80 and 90 °C)		$\beta_1 = -5.88 \bullet 10^3$		
				$\beta_2 = 47.7$		
3	$PS = \beta_0 + \beta_1 pH + \beta_2 pH^2$	80 °C	6	$\beta_0 = 1.97 \bullet 10^3$	0.881	3.8
				$\beta_1 = -526$		
				$\beta_2 = 37.4$		
		90 °C	6	$\beta_0 = 7.87 \bullet 10^3$	0.992	3.2
				$\beta_1 = -2.23 \bullet 10^3$		
				$\beta_2 = 161$		
4	$PS = \beta_0 + \beta_1 pH + \beta_2 pH^2 + \beta_3 T$	Integrated	14	$\beta_0 = 4.78 \bullet 10^3$	0.791	11.7
		(80 and 90 °C)		$\beta_1 = -1.38 \bullet 10^3$		
				$\beta_2 = 99.1$		
				$\beta_3 = 1.60$		

N = 18; DF err, degrees of freedom for error; coef.: coefficients; β_{0-3}, prediction coefficients; LB, light backscatter intensity (bits); PS, particle size z-average (nm); R^2, determination coefficient; SEP, standard error of prediction for the model (bits for LB and nm for PS).
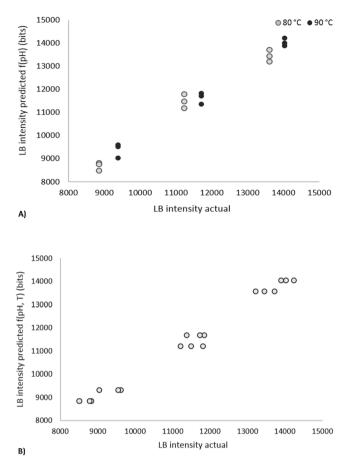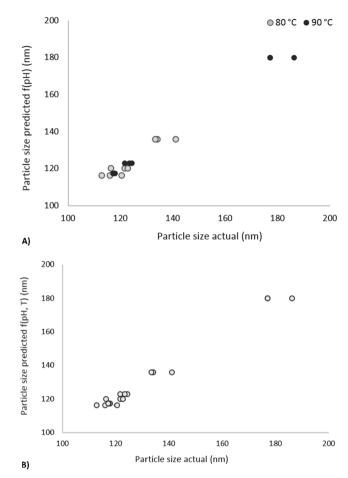
Fig. 2. (A) Model 1: Light backscatter maximum intensity modeled as a function of pH at 80 and 90 °C (N = 18); (B) Model 2: Light backscatter maximum intensity modeled as a function of pH and temperature (N = 18).

Fig. 3. (A) Model 3: Particle size z-average modeled as a function of pH at 80 and 90 °C (N = 18); (B) Model 4: Particle size z-average modeled as a function of pH and temperature (N = 18).
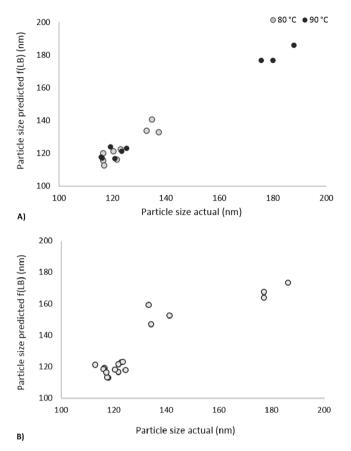
measurements in milk as it is inexpensive and nondestructive, and can easily be implemented for in-line application. Initial work by Anema and Li (2003) presented a strong correlation of denatured whey proteins attached to the casein micelle and changes in particle size. As well, previous work has shown good correlation between both the light backscatter maximum intensity (r = 0.77) (Taterka and Castillo, 2015) and bound whey protein (r = 0.70) (Taterka and Castillo, 2018) to changes in casein micelle particle size. A model of particle size as a function of light backscatter intensity would enable in-line monitorization of the extent of changes in denaturation in heat treated milk. Since in-line process control shows promise toward optimization of the manufacturing process of milk and milk products such as cheese and yogurt, a simple, non-invasive, quick and inexpensive technique providing real time information about the binding of denatured whey proteins to the casein micelle and/or soluble aggregate formation would be highly advantageous to the dairy industry. Note that different processing factors inducing particle interactions in milk are relevant for milk functionality and would be associated to changes in particle size that might be real-time revealed using light backscatter measurements. And a relatively simple equation for the prediction of particle size as a function of the light backscatter maximum, was developed:

$$PS = \beta_0 + \beta_1 LB + \beta_2 LB^2 \tag{5}$$

Model 5 (Table 2) used a quadratic form and yielded high R^2 values in the case of 80 and 90 °C models ($R^2 = 0.847$ and 0.992, respectively) (Fig. 4A) and temperature "combined" models ($R^2 = 0.825$) (Fig. 4B). As it can be observed in Table 1 and 2, temperature separate models had higher R^2 and lower standard error of prediction (SEP) in nearly all models (except in the case of the temperature "integrated" Model 2,

Table 2
Predictive model 5 for particle size z-average (PS).

Model	Prediction equation	Temperature	DF err	Regression coef.	R^2	SEP
5	$PS = \beta_0 + \beta_1 LB + \beta_2 LB^2$	80 °C	6	$\beta_0 = 219$	0.847	4.4
				$\beta_1 = -0.0221$		
				$\beta_2 = 1.18 \cdot 10^{-6}$		
		90 °C	6	$\beta_0 = 572$	0.992	3.1
				$\beta_1 = -0.0899$		
				$\beta_2 = 4.41 \cdot 10^{-6}$		
		Combined (80 and 90 °C)	15	$\beta_0 = 479$	0.825	10.3
				$\beta_1 = -0.0725$		
				$\beta_2 = 3.59 \cdot 10^{-6}$		

N = 18; DF err, degrees of freedom for error; coef.: coefficients; β_{0-3}, prediction coefficients; *LB*, light backscatter intensity (bits); *PS*, particle size z-average (nm); R^2, determination coefficient; SEP, standard error of prediction for the model (bits for *LB* and nm for *PS*).

which has a higher R^2 and lower SEP value than Model 1 temperature separated at 80 °C). However, all temperature "integrated" models maintained an R^2 value higher than or equal to 0.791 and all 80 and 90 °C models with R^2 greater than 0.847 (Tables 1 and 2).

Other authors have looked at various characteristics of milk using NIR (780–2500 nm) or mid-infrared (MIR) (2500–15000 nm)

Fig. 4. (A) Model 5 temperature separated: Particle size z-average modeled as a function of light backscatter maximum intensity (N = 18); (B) Model 5 (temperature combined): Particle size z-average modeled as a function of light backscatter maximum intensity (N = 18).

spectroscopy techniques (Iñón et al., 2004; Wu et al., 2011) by incorporating a combination of chemometric techniques for analyzing the range of spectral data in order to form useful prediction models. The NIR region is widely used in milk analysis and therefore the spectral regions which correspond to moisture content, milk fat, protein, lactose and other milk components are well-characterized (Kamishikiryo-Yamashita et al., 1994; Laporte and Paquin, 1999; Robert et al., 1987; Tsenkova et al., 1999). However, it should be noted that although the light backscatter spectral range (200–1100 nm) has been studied with respect to milk particulate components, such as the fat globule and casein micelle (Castillo et al., 2005a,b; Fagan et al., 2008; Lamb et al., 2013), regions that correspond to specific characteristic and or physicochemical changes in milk have yet to be fully characterized.

3.2. Waveband ratio predictive models

3.2.1. Waveband ratio selection

In the present study, a good modelization for particle size as a function of light backscatter maximum intensity (model 5) has been observed. However Lamb et al. (2013) found improved predictions when using an alternative technique of waveband ratios for modeling whey protein denaturation as a function of light backscatter intensity. Lamb et al. (2013) defined waveband ratios as the average intensity at each 25 nm portion of the light backscatter spectra, and created waveband ratio combinations, which were used to form predictive models. Implementing a technique of ratios, or a combination of parameters, has been used successfully in prediction models found in the literature (Castillo et al., 2000; M. Castillo et al., 2005a,b; Fagan et al., 2008), which in certain cases was found to improve the accuracy of modeling.

In addition, a technique of ratios has been used to decrease the number of variables used in the model by combining certain parameters, such as in the case of Fagan et al. (2008), which incorporated a non-optical milk-fat protein ratio into the model for curd moisture content. Therefore, in order to pursue a higher level of accuracy for predictive models, a method of waveband ratios was also tested. Note that Fagan et al. (2008) tested Partial Least Squares (PLS) regression analysis for use in optical sensor development and found little improvement in models when compared to individual wavelength and waveband ratio models. In addition, complex model development techniques are inherently more complicated and expensive as they require sensors that use multiple wavelength analysis and, as a result, it was opted out of using any more than two wavelength combinations for the development of models in the present study.

Initially, wavebands of 15, 25 and 35 nm portions of the light backscatter spectra (from 200 to 1100 nm) were investigated in preliminary models, however no major differences were observed. As a result, wavebands were defined as an average of the intensity for each 35 nm portion of the spectra. Thus, a total of 27 waveband regions were obtained. The wavebands were then combined, using each waveband combination in both the numerator and denominator, to come up with all possible ratios. Predictive models were determined using the preselected waveband ratios that were found to best correspond to changes in particle size as a function of pH (i.e., assuming an exponential function). Table 3 summarizes the top 10 ratios, which exhibited the highest R^2 values used for the model development and based on the total R^2 value from a sum of the four models which used waveband ratios (Models 6 (80 °C), 6 (90 °C), 7 and 10). Fig. 5 illustrates the typical light backscatter profile obtained from 200 to 1100 nm scans. Highlighted are the two sections of wavebands which were used to obtain the best models. Wavebands 6 and 7 (range of 388–458 nm) were selected as numerator values that corresponded with models showing the highest R^2 values. The number of denominators that corresponded to the highest R^2 values for the models was larger, wavebands 15–20 (range of 703–878 nm). All ratios in the top 10 models fell within the previously described numerator and denominator ranges. Numerator values that produced the 10 highest R^2 represent a region with relatively little change among the different samples at various experimental conditions (pH and temperature of heat treatment), whereas denominator values tended to lie in a region with more notable changes (Fig. 5). The technique of using numerator and denominator values presented as a ratio and/or as individual values was further detailed in Section 2.

3.2.2. Models

Using the ratio ($R_{6,18}$) which yielded the highest R^2 for particle size, prediction models were developed for particle size (z-average) and light backscatter waveband ratio ($R_{6,18}$) as a function of pH in temperature "separated" and temperature "integrated" models (Table 4). In

Table 3
R^2 top 10 ratios for modeling particle size.

ratio	numerator	denominator	R^2 PS f(LB ratio)	R^2_{avg} PS f(LB ratio)	R^2 LB ratio f(pH,T)	R^2_{avg} LB ratio f(pH,T)
6_18	388	808	0.979	0.952	0.991	0.993
6_17	388	773	0.979	0.951	0.992	0.993
6_16	388	843	0.978	0.949	0.993	0.993
7_16	388	738	0.977	0.949	0.992	0.993
7_15	458	773	0.977	0.948	0.992	0.993
6_19	458	738	0.980	0.954	0.987	0.989
7_17	458	703	0.978	0.951	0.989	0.991
6_15	458	808	0.976	0.948	0.991	0.992
6_20	388	878	0.981	0.955	0.982	0.985
7_18	388	703	0.976	0.951	0.986	0.989

R^2, determination coefficient; R^2_{avg}, average determination coefficient of 80 and 90 °C; LB, light backscatter intensity (bits); T, temperature of heat treatment (°C); PS, particle size z-average (nm); numerator and denominator (nm).
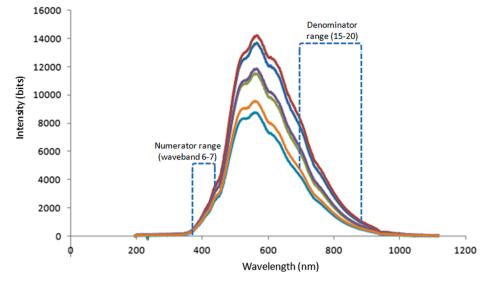
Fig. 5. Light backscatter spectra showing selected numerator and denominator regions used for modeling.

Table 4

Models using top 1 ratio $R_{6,18}$ for prediction of $\frac{I_n}{I_d}$ and an exponential factor for prediction of particle size z-average (PS).

Model	Prediction equation	Temperature	DF err	Regression coef.	R^2	SEP
6	$\frac{I_n}{I_d} = \beta_o + \beta_1 pH$	80 °C	7	$\beta_0 = 0.454$	0.991	$3.69 \bullet 10^{-3}$
				$\beta_1 = -0.103$		
		90 °C	7	$\beta_0 = 0.488$	0.996	$2.57 \bullet 10^{-3}$
				$\beta_1 = -0.106$		
7	$\frac{I_n}{I_d} = \beta_o + \beta_1 pH + \beta_2 T$	Integrated	15	$\beta_0 = 0.383$	0.993	0.0377
		(80 and 90 °C)		$\beta_1 = -0.104$		
				$\beta_2 = -1.03 \bullet 10^{-3}$		
8	$PS = \beta_o + e^{(\alpha_0 + \alpha_1 pH)}$	80 °C	24	$\beta_0 = 115$	0.871	3.49
				$\alpha_0 = -25.2$		
				$\alpha_1 = 3.52$		
		90 °C	21	$\beta_0 = 117$	0.988	3.09
				$\beta_1 = -41.2$		
				$\beta_2 = 5.87$		
9	$PS = \beta_o + e^{(\alpha_0 + \alpha_1 pH + \alpha_2 T)}$	Integrated	47	$\beta_0 = 117$	0.976	3.32
		(80 and 90 °C)		$\alpha_0 = -29.7$		
				$\alpha_1 = 5.78$		
				$\alpha_2 = 0.121$		

N = 18 for I_n/I_d models, N = 51 for PS models; DF err, degrees of freedom for error; coef.: coefficients; β_{0-2} and α_{0-2}, prediction coefficients; R^2, determination coefficient; SEP, standard error of prediction for the model (dimensionless for $\frac{I_n}{I_d}$ and nm for PS).

comparison to models that were developed to determine $R_{6,18}$ (models 6–7; 6, 7) (Fig. 6), models for z-average (models 8–9; 8, 9) (Fig. 7) incorporated an exponential factor into equations. As mentioned previously and visualized in Fig. 1B, particle size z-average tends to follow a more exponential curve with respect to pH when compared to light backscatter maximum intensity, which follows a linear trend with pH (Fig. 1A). Equations for predictions are as follows:

$$\frac{I_n}{I_d} = \beta_0 + \beta_1 pH \tag{6}$$

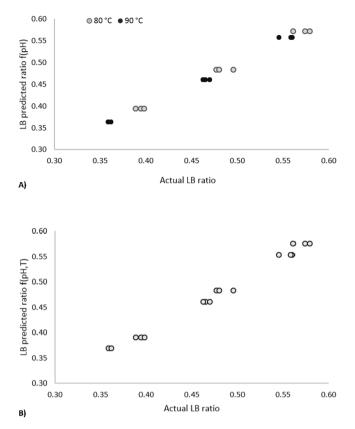$$\frac{I_n}{I_d} = \beta_0 + \beta_1 pH + \beta_2 T \tag{7}$$

$$PS = \beta_0 + e^{(\alpha_0 + \alpha_1 pH)} \tag{8}$$

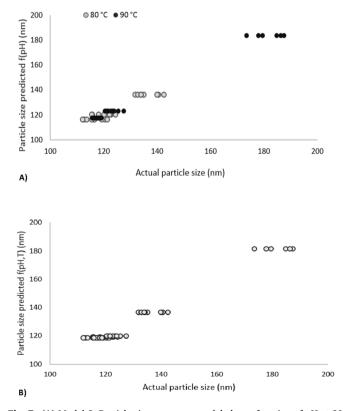$$PS = \beta_0 + e^{(\alpha_0 + \alpha_1 pH + \alpha_2 T)} \tag{9}$$

where I_n is the light backscatter intensity at the pre-selected numerator, I_d is the light backscatter intensity at the pre-selected denominator, PS is the particle size z-average, T is the heat treatment temperature, β_{0-2} and α_{0-2} are prediction coefficients.

A previous work has shown that changes in light backscatter signal and particle size have been found to be highly correlated to pH (Taterka and Castillo, 2015). Thus in model development, it is reasonable that pH has been found to be a critical factor. Castillo et al. (2006) developed models for cutting time prediction in cottage cheese manufacture also using pH as a variable for predictions, whereas reasonably accurate prediction models using pH as a variable have been developed by Toffanin, De Marchi, Lopez-Villalobos, and Cassandro (2015) in order to determine characteristics of milk quality and coagulation properties using MIR analysis. As pH is a useful tool in the present study for the determination of particle size and light backscatter intensity, it may be complementary to utilize a simple technique of milk pH measurement for the determination of changes in the milk matrix with heat treatment. Unfortunately, simple laboratory pH meters generally contain glass components and other reagents that are not approved for in-line use in

Fig. 6. (A) Model 6: Light backscatter ratio R6,18 (dimensionless) modeled as a function of pH (N = 18); (B) Model 7: Light backscatter ratio R6,18 (dimensionless) modeled as a function of pH and temperature (N = 18).

Fig. 7. (A) Model 8: Particle size z-average modeled as a function of pH at 80 and 90 °C (N = 51); (B) Model 9: Particle size z-average modeled as a function of pH and temperature (N = 51).

the food industry (Wesstrom, 1992). A preliminary work by Arango et al. (2020) has exhibited that optical sensors are useful for the replacement of pH meters during acid-coagulation of milk in which pH predictions as a function of light scatter techniques were developed successfully with great accuracy (R^2 greater than 0.99). Thus, the incorporation of an optical technology for the determination of pH-specific denaturation mechanisms may be of interest for investigation in the present study. As a result, models for particle size using the light backscatter spectrum have been developed with good correlations. Utilizing the ratio values showed to greatly improve R^2 when comparing model 5, which did not use ratio values, (Table 2) (Fig. 4) to model 10, which did (Table 5) (Fig. 8). Model 10 is presented below, utilizing $R_{6,18}$ in an exponential type equation to model particle size changes.

$$PS = \beta_0 + e^{(\alpha_0 + \alpha_2 \frac{I_n}{I_d})} \tag{10}$$

An interesting finding in prediction models for particle size z-average (Model 8–10) is that the coefficient β_0 lies within the range of the initial casein micelle particle radius (~112–120 nm, Tables 4, 5) (Figs. 7, 8). The remainder of the prediction equation includes an exponential factor of other predictors (8–10), which for Model 10 corresponds to a wave-band ratio.

Since binding has been found to be the main contributor in the increase in particle size (Taterka and Castillo, 2015), it may be suggested that the binding reaction follows an exponential growth curve. Thus, using the average initial casein micelle particle size plus some exponential increase as a result of attachment and/or aggregation, it may be possible to predict casein micelle particle size growth and, in turn, denaturation as a function of particle size increase. In the present study, the average radius was 119.7 nm using reconstituted milk from the same lot, comparable to average diameters reported in Martin, Williams, and Dunstan (2007) which saw an average of 231.0 ± 1.6 nm in reconstituted milk stirred for 35 min. However, it should be noted that casein micelle average size varies significantly depending on batch and thus this hypothesis should be tested using different milk batches.

3.3. Comparison of single wavelength and ratio models

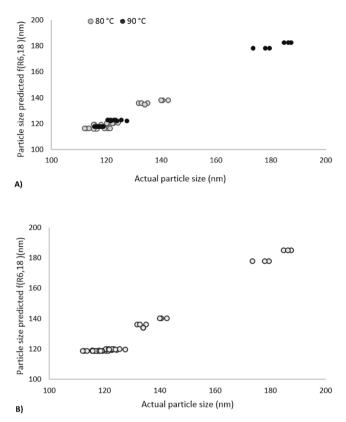#### 3.3.1. Comparison of models

Two sets of models have been developed and discussed: models developed using the maximum light backscatter intensity (*LB*) (Tables 1, 2) and models using ratio $R_{6,18}$ and an exponential factor in z-average models (Table 4, 5). Models from Table 1 are complimentary to models from Table 4 as such: Model 1 and 6, 2 and 7, 3 and 8, 4 and 9; and 5 and 10 in Table 2 and Table 5, respectively. In most cases, using a ratio value and/or adding an exponential factor to particle size prediction models acted to increase the R^2 value of the model, except model 3 compared to

Table 5
Predictive Model 10 for particle size z-average (PS).

Model	Prediction equation	Temperature	DF err	Regression coef.	R^2	SEP
10	$PS = \beta_0 +$ $e^{(\alpha_0 + \alpha_2 \frac{I_n}{I_d})}$	80 °C	24	$\beta_0 = 115$	0.911	2.89
				$\alpha_0 = $ -9.22		
				$\alpha_1 = 15.6$		
		90 °C	21	$\beta_0 = 117$	0.993	2.36
				$\alpha_0 = $ -12.5		
				$\alpha_1 = 23.2$		
		Combined (80 and 90 °C)	48	$\beta_0 = 119$	0.979	3.04
				$\alpha_0 = $ -17.5		
				$\alpha_1 = 37.0$		

N = 51; DF err, degrees of freedom for error; coef.: coefficients; β_{0-2} and α_{0-2}, prediction coefficients; *PS*, particle size z-average (nm); R^2, determination coefficient; SEP, standard error of prediction for the model (nm).

Fig. 8. (A) Model 10: Particle size z-average modeled as a function of light backscatter ratio R6,18 at 80 and 90 °C (N = 51); (B) Model 10 (temperature combined): Particle size z-average modeled as a function of light backscatter ratio R6,18 (N = 51).

model 8 where adding an exponential factor gave a slightly lower R^2 for both 80 (model 3: 0.881, model 8: 0.871) and 90 °C models (model 3: 0.992, model 8: 0.988). This may be due to the fact that there is a large increase from 80 °C to 90 °C in particle size at pH 6.3. This difference makes it necessary for temperature "integrated" models to include an exponential factor ($R^2 = 0.791$ -Table 1- versus 0.976 -Table 4- with exponential factor), whereas temperature "separated" models using a simpler equation still maintain a high R^2 value for the models ($R^2 = 0.881$ and 0.992 -Table 1- versus 0.871 and 0.988 -Table 4-).

For the modelization of particle size z-average as a function of light backscatter intensity, model 5 and model 10 use light backscatter intensity values (model 5: LB, model 10: $R_{6,18}$). Both models show high R^2 values, however it can be seen that the addition of an exponential factor, as well as using $R_{6,18}$ in model 10, acts to increase R^2 in both temperature "separate" and "combined" models (R^2 for 80 °C; 90 °C and "combined" models were 0.847, 0.992, 0.825 for model 5: and 0.911, 0.993, 0.979 for model 10, respectively). This effect is also observed in the case of the temperature "integrated" models (Models 4 and 9) in which the exponential addition greatly increases the R^2 (0.791 and 0.976, respectively). On the other hand, when comparing Model 3 with Model 8, the simple quadratic form yields a prediction with a higher R^2 (80 °C: 0.881, 90 °C: 0.992) than when the exponential form is added to the equation (80 °C: 0.871, 90 °C: 0.988). In this case, models that are produced at each respective temperature separately, using a linear model, supports each prediction sufficiently and to a better extent than the exponential model. Nonetheless, there is a good correlation between actual and predicted values in all models, thus the light backscatter technique shows promise toward the prediction of particle size changes as a function of milk pH and temperature treatment. Since changes in particle size have been mainly found to be a result of the binding reaction of denatured whey proteins attaching to the surface of the casein

micelle, this gives some insight into aspects of whey protein denaturation in milk.

3.3.2. Analysis based on various portions of R6,18

Since two techniques were implemented for modeling, one using a single intensity at the maximum value (LB; in this section represented by I_m), and the other using a ratio of intensities ($R_{6,18}$), which acted to improve the models in most cases, it was of interest to test whether using only the numerator or denominator of the ratio $R_{6,18}$ (I_n or I_d, respectively) may also yield good predictions individually. Note that a single wavelength prediction is always more convenient for industrial implementation of a sensor technology. Thus, in order to confirm the need for ratios in the development of prediction models, model 5 was tested using the maximum intensity (I_m), the intensity used in the numerator (I_n) of $R_{6,18}$, and the intensity of the denominator (I_d) used in $R_{6,18}$. As summarized in Table 6, it can be seen that both I_m ($R^2 = 0.847$ and 0.992) and I_d ($R^2 = 0.779$ and 0.986) showed good R^2 for both 80 and 90 °C models, respectively. On the other hand, I_n yielded less reliable R^2 values ($R^2 = 0.465$ and 0.773). In temperature "combined" models I_d was the model with the highest R^2 (0.890) and lowest SEP (8.20 nm), compared to I_m and I_n models ($R^2 = 0.825$; 0.491, SEP = 10.3; 17.6 nm, respectively). As I_n, in general, shows poor correlation of predicted and actual values in models, it leads us to believe that it may not be necessary for

Table 6
Models using individual wavelength values to model particle size as a function of the maximum intensity value ($I_m = 570$ nm), the intensity used in the numerator (I_n) and the intensity of the denominator (I_d).

Model	Prediction equation	Temperature	DF err	Regression coef.	R^2	SEP
5	$PS = \beta_0 + \beta_1 I_m + \beta_2 I_m^2$	80 °C	6	$\beta_0 = 1.18\bullet 10^{-6}$ $\beta_1 = 0.0221$ $\beta_2 = 219$	0.847	4.36
		90 °C	6	$\beta_0 = 4.41\bullet 10^{-6}$ $\beta_1 = 0.0899$ $\beta_2 = 572$	0.992	3.12
		Combined (80 and 90 °C)	15	$\beta_0 = 479$ $\beta_1 = -0.0725$ $\beta_2 = 3.59\bullet10^{-6}$	0.825	10.3
11	$PS = \beta_0 + \beta_1 I_n + \beta_2 I_n^2$	80 °C	6	$\beta_0 = -2.44\bullet 10^{-5}$ $\beta_1 = -0.110$ $\beta_2 = 49.9$	0.465	8.14
		90 °C	6	$\beta_0 = -1.36\bullet 10^{-3}$ $\beta_1 = -1.99$ $\beta_2 = 847$	0.773	16.6
		Combined (80 and 90 °C)	15	$\beta_0 = 436$ $\beta_1 = -0.920$ $\beta_2 = 6.60\bullet 10^{-4}$	0.491	17.6
12	$PS = \beta_0 + \beta_1 I_d + \beta_2 I_d^2$	80 °C	6	$\beta_0 = 4.18\bullet 10^{-6}$ $\beta_1 = 0.0231$ $\beta_2 = 148$	0.889	3.72
		90 °C	6	$\beta_0 = 1.36\bullet 10^{-5}$ $\beta_1 = 0.0879$ $\beta_2 = 258$	0.994	2.60
		Combined (80 and 90 °C)	15	$\beta_0 = 216$ $\beta_1 = 4.46\bullet 10^{-5}$ $\beta_2 = -0.134$	0.890	8.20

N = 18, DF err, degrees of freedom for error; $\beta_{0\text{-}2}$, prediction coefficients; R^2, determination coefficient; SEP, standard error of prediction for the model (nm); I_m: single intensity at the maximum value; I_n: intensity used in the numerator; I_d: intensity of the denominator.

model development. A possible reason for this will be introduced below. In fact, I_d yields even higher R^2 values than I_m; possibly showing more potential in the corresponding portion of the spectrum (808 nm) for the determination of various changes in milk with heat treatment and varying milk pH as compared with the peak of maximum intensity (570 nm). Previous work in the literature (Lamb et al., 2013) found it useful to form ratios for predictive models however in the present study the regions of I_m and I_d used individually appear to be much more important than I_n, at least in the case of quadratic models for particle size as a function of light backscatter intensity. Models used for comparison (Table 6) do not include an exponential factor, as it was not possible to model individual intensity values without modifying the exponential equation. Since little change in intensity occurs between sample treatments in the regions where numerators with the highest R^2 for models occur (Table 3), it is suggested that the portion of the spectrum corresponding to the denominators (703–878 nm) is in fact representing the particle size information while numerator values act as a normalization factor for models. Additionally, exponential models tend to yield high R^2 values and include the coefficient β_0, which may be representative of the initial particle size ratio values. Considering these two observations, it is proposed that development of an in-line sensor using only one wavelength plus some predetermined normalization factor (for example, replacing the numerator with a constant) implemented into an exponential equation should yield a considerably accurate model, and in addition, would be less costly and complicated than a dual wavelength optical sensor.

4. Conclusion

Two prediction models were developed to estimate particle size z-average only as a function of the intensities of a single wavelength/waveband combinations on the light backscatter spectra (Model 5 -single wavelength- and 10 -wavelength ratio-), both showing good correlation between actual and predicted values. However, improvements were found with the incorporation of $R_{6,18}$ and the addition of an exponential factor (Model 10) compared to the quadratic model using LB (i.e., I_m) (Model 5). As changes in particle size diameter have been found to be primarily a result of the attachment of denatured whey proteins to the surface of the casein micelle, this model provides useful information regarding the potential modelization of WP denaturation. Taking into account the investigated models, the suggested model for particle size as a function of light backscatter intensity would include both an exponential component, an initial intercept which corresponds to an approximate value of initial particle size and a light backscatter normalized value or a waveband ratio. Further work may give more insight into a more improved model. Nonetheless, this experiment yields useful preliminary information toward the development a comprehensive model for the determination of whey protein denaturation with the potential for implementation of an in-line optical sensor for in plant processed milk.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Anema, S. G. (1997). The effect of chymosin on κ-casein-coated polystyrene latex particles and bovine casein micelles. *International Dairy Journal, 7*(8), 553–558.

Anema, S. G., Lowe, E. K., & Li, Y. (2004). Effect of pH on the viscosity of heated reconstituted skim milk. *International Dairy Journal, 14*(6), 541–548.

Anema, S. G., & Li, Y. (2003). Association of denatured whey proteins with casein micelles in heated reconstituted skim milk and its effect on casein micelle size. *Journal of Dairy Research, 70*, 73–83.

Arango, O., & Castillo, M. (2018). A method for the inline measurement of milk gel firmness using an optical sensor. *Journal of Dairy Science, 101*, 3910–3917.

Arango, O., Trujillo, A. J., & Castillo, M. (2020). Inline control of yoghurt fermentation process using a near infrared light backscatter sensor. *Journal of Food Engineering, 277*, Article 109885.

Castillo, M., Fagan, C.C., Payne, F.A., ÓDonnell, C.P., ÓCallaghan, D.J., 2007. Optical measurement of kinetic changes in curd moisture content and whey fat concentration during syneresis in cheese manufacturing. J. of Dairy Sci. 90(Supple. 1), 272.

Castillo, M., Payne, F. A., Hicks, C., & Lopez, M. (2000). Predicting cutting and clotting time of coagulating goat's milk using diffuse reflectance: Effect of pH, temperature and enzyme concentration. *International Dairy Journal, 10*, 551–562.

Castillo, M., Payne, F. A., & Lopez, M. (2005a). Preliminary evaluation of an optical method for modeling the dilution of fat globules in whey during syneresis of cheese curd. *Applied Engineering in Agriculture, 21*(2), 265–269.

Castillo, M., Payne, F. A., López, M. B., Ferrandini, E., & Laencina, J. (2005b). Optical sensor technology for measuring whey fat concentration in cheese making. *Journal of Food Engineering, 71*, 354–360.

Castillo, M., Payne, F. A., & Shea, A. P. (2005c). Development of a combined sensor technology for monitoring coagulation and syneresis operations in cheese making. *Journal of Dairy Science, 88(Suppl. 1):142*, (Abstr.).

Castillo, M., Payne, F. A., Wang, T., & Lucey, J. (2006). Effect of temperature and inoculum concentration on prediction of both gelation time and cutting time. *Cottage cheese-type gels. Int. dairy J., 16*, 147–152.

Diaz-Carrillo, E., Muñoz-Serrano, A., Alonso-Moraga, A., & Serradilla-Manrique, J. M. (1993). Near infrared calibrations for goat's milk components: Protein, total casein, αs-, β-and k-caseins, fat and lactose. *Journal of Near Infrared Spectroscopy, 1*, 141–146.

Donato, L., Guyomarc'h, F., 2009. Formation and properties of the whey protein/κ-casein complexes in heated skim milk – A review. Dairy Sci. Technol. 89, 3–29.

Downey, G., Robert, P., Bertrand, D., & Kelly, P. M. (1990). Classification of Commercial Skim Milk Powders According to Heat Treatment Using Factorial Discriminant Analysis of Near-Infrared Reflectance Spectra. *Applied Spectroscopy, 44*, 150–155.

Fagan, C. C., Castillo, M., O'Donnell, C. P., O'Callaghan, D. J., & Payne, F. A. (2008). On-line prediction of cheese making indices using backscatter of near infrared light. *International Dairy Journal, 18*, 120–128.

Fagan, C. C., Leedy, M., Castillo, M., Payne, F. A., O'Donnell, C. P., & O'Callaghan, D. J. (2007). Development of a light scatter sensor technology for on-line monitoring of milk coagulation and whey separation. *Journal of Food Engineering, 83*, 61–67.

García, J., 2004. Aplicaciones de la tecnología NIRS en la industria agroalimentaria. Nuevas Tecnologías para el Control proceso y Prodroducto en la Industria Alimentaria. Ed. Secretariado de Publicaciones en Intercambio Editorial de la Universidad de Valladolid. Valladolid.

Giangiacomo, R., Braga, F., & Galliena, C. (1991). Use of near-infrared spectroscopy to detect whey powder mixed with milk powder. *Near Infrared Spectroscopy, Making Light Work., 399–407*.

Iñón, F. A., Garrigues, S., & de la Guardia, M. (2004). Nutritional parameters of commercially available milk samples by FTIR and chemometric techniques. *Analytica Chimica Acta, 513*, 401–412.

Kamishikiryo-Yamashita, H., Oritani, Y., Takamura, H., & Matoba, T. (1994). Protein content in milk by near-infrared spectroscopy. *Journal of Food Science, 59*(2), 313–315.

Kethireddipalli, P., Hill, A. R., & Dalgleish, D. G. (2010). Protein interactions in heat-treated milk and effect on rennet coagulation. *International Dairy Journal, 20*(12), 838–843.

Lamb, A., Payne, F., Xiong, Y. L., & Castillo, M. (2013). Optical backscatter method for determining thermal denaturation of β-lactoglobulin and other whey proteins in milk. *Journal of Dairy Science, 96*, 1356–1365.

Laporte, M.-F., & Paquin, P. (1999). Near-Infrared Analysis of Fat, Protein, and Casein in Cow's Milk. *Journal of Agriculture and Food Chemistry, 47*, 2600–2605.

Martin, G. J. O., Williams, R. P. W., & Dunstan, D. E. (2007). Comparison of casein micelles in raw and reconstituted skim milk. *Journal of Dairy Science, 90*, 4543–4551.

Pedretti, N., Bertrand, D., Semenou, M., Robert, P., & Giangiacomo, R. (1993). Application of an experimental design to the detection of foreign substances in milk. *Journal of Near Infrared Spectroscopy, 1*, 174.

Robert, P., Bertrand, D., Devaux, M. F., & Grappin, R. (1987). Multivariate analysis applied to near-infrared spectra of milk. *Analytical Chemistry, 59*, 2187–2191.

Sørensen, L., & Jepsen, R. (1997). Detection of cheese batches exposed to Clostridium tyrobutyricum spoilage by near infrared spectroscopy. *Journal of Near Infrared Spectroscopy, 5*, 91.

Taterka, H., & Castillo, M. (2015). The effect of whey protein denaturation on light backscatter and particle size of the casein micelle as a function of pH and heat-treatment temperature. *International Dairy Journal, 48*, 53–59.

Taterka, H., & Castillo, M. (2018). Analysis of the preferential mechanisms of denaturation of whey protein variants as a function of temperature and pH for the development of an optical sensor. *International Journal of Dairy Technology, 71*(1), 226–235.

Toffanin, V., De Marchi, M., Lopez-Villalobos, N., & Cassandro, M. (2015). Effectiveness of mid-infrared spectroscopy for prediction of the contents of calcium and phosphorus, and titratable acidity of milk and their relationship with milk quality and coagulation properties. *International Dairy Journal, 41*, 68–73.

Tsenkova, R., Atanassova, S., Toyoda, K., Ozaki, Y., Itoh, K., & Fearn, T. (1999). Near-infrared spectroscopy for dairy management: Measurement of unhomogenized milk composition. *Journal of Dairy Science, 82*, 2344–2351.

Vasbinder, A. J., & de Kruif, C. G. (2003). Casein–whey protein interactions in heated milk: The influence of pH. *International Dairy Journal, 13*(8), 669–677.

Wesstrom, O., 1992. Inline pH Measurement for the Food / Dairy and Beverage Industry.

Wu, D., Nie, P., He, Y., & Bao, Y. (2011). Determination of Calcium Content in Powdered Milk Using Near and Mid-Infrared Spectroscopy with Variable Selection and Chemometrics. *Food and Bioprocess Technology, 5*, 1402–1410.