Phenotypic and Genetic Variation of Milk Fat Components Incorporating Mid-Infrared Technology

by

Allison L. Fleming

A Thesis
presented to
The University of Guelph

In partial fulfilment of requirements
for the degree of
Doctor of Philosophy
in
Animal and Poultry Science

Guelph, Ontario, Canada

© Allison L. Fleming, August, 2016
Milk fat influences the economic, nutritional, flavour, and physico-chemical properties of milk and milk products. Genetic selection for fat components is therefore of interest. This would require an efficient means of phenotyping these traits routinely, on a large scale. The purpose of this study was to assess the ability to predict milk fat components using mid-infrared (MIR) spectroscopy and to examine the phenotypic and genetic variation of these traits.

Individual cow milk samples from multiple breeds and herds were collected during routine milk recording multiple times throughout the lactation and the MIR spectra obtained. Milk samples were further analyzed for average milk fat globule (MFG) size and fatty acid profile using gold standard methodologies. Partial least squares regression models were employed to develop equations to predict the milk fat component traits from the MIR spectra of milk samples. Genetic parameters were estimated in Holsteins for both the measured and predicted MFG size and predicted fatty acids, along with their genetic correlations with milk production traits using multi-trait animal models.

Milk fatty acid contents could be predicted with variable accuracy depending on the calibration set used and the concentration of the fatty acid in milk. In first-parity Holsteins, the predicted groups of fatty acids were found to have genetic correlations of similar magnitude with
already recorded milk production traits. Differences in the genetic correlations with fat yield, fat and protein percentages, and fat to protein ratio were found for different fatty acid groups, including different trends in average daily genetic correlations at the beginning of lactation. Average MFG size was significantly affected by herd, breed, days in milk, season, and milking period. The prediction of MFG size from MIR spectra was poor, and predicted values had greater phenotypic correlations with fat percentage than their measured counterparts. A moderate heritability was found for average MFG size, but predicted values had very strong genetic correlations with fat percentage, limiting their utility. Mid-infrared technology provides an opportunity to obtain novel trait phenotypes, including fatty acids, for all milk recorded cows that could be used to help advance the Canadian dairy industry.
ACKNOWLEDGEMENTS

First and foremost I would like to thank my advisors, Dr. Filippo Miglior and Dr. Flavio Schenkel, who perfectly complemented one another and cultivated my growth through this time. Filippo, you were a constant source of guidance and support for me. Thank you for your patience and understanding, keeping me focused and on track, and generously sharing your expertise. I was extremely fortunate to have an advisor who cared so much about both my work and my development. Flavio, your immense knowledge and attention to detail fostered my deeper understanding of animal breeding and critical thinking.

I am indebted to my advisory committee members, Drs. Ayesha Ali, Milena Corredig, Bonnie Mallard, and Mehdi Sargolzaei, a multidisciplinary group who brought unique perspectives and insights, broadened my learning experience, and gave me the capacity to connect ideas. The discussion and recommendations provided by this group were invaluable to the completion of this work. I would also like to thank Dr. John Cole and Dr. Christine Baes for participating in my final exam.

I am grateful to everyone involved in CGIL and the Department of Animal Biosciences for making my time as a graduate student enjoyable. I was able to learn so much from the many people who were always willing to share their knowledge. I was lucky to be a part of an amazing and supportive group. A special thank you to Dr. Francesca Malchiodi who helped me to navigate my way as a PhD student, and became my go-to person for all situations. Thank you as well to Dr. Astrid Koeck for her assistance with data preparation and analysis.

This project would not have been possible without all of the hard work of Jin Chen and colleagues in the Department of Food Science. Their work in the lab analysis and organization of the vast number of milk samples they received provided me great data to work with. I would like to acknowledge all of the dairy producers participating in this project. I am also appreciative to CanWest DHI and Valacta for their work in the selection of herds and milk sample collection.

Finally, a very sincere thank you to my family and friends who supported my pursuit of this degree, and who took a strong interest in everything I was doing, even if they did not understand most of it.
# TABLE OF CONTENTS

1. GENERAL INTRODUCTION ................................................................................................ 1
   INTRODUCTION .............................................................................................................. 1
   OBJECTIVES ..................................................................................................................... 3

2. LITERATURE REVIEW ........................................................................................................ 5
   MILK LIPIDS ..................................................................................................................... 5
   MILK FATTY ACIDS ....................................................................................................... 5
   Origins of Milk Fatty Acids ....................................................................................... 5
   Potential Health Effects ............................................................................................... 6
   Technological and Sensory Properties ....................................................................... 8
   Factors Influencing Fatty Acid Composition ............................................................. 8
   Fatty Acid Profile as an Indicator of Other Traits ....................................................... 11
   MILK FAT GLOBULES .................................................................................................. 14
   Potential Health Effects of the Milk Fat Globule Membrane ....................................... 15
   MILK FAT GLOBULE SIZE ........................................................................................... 18
   Determining and Expressing Milk Fat Globule Size .................................................... 18
   Factors Influencing Milk Fat Globule Size .................................................................... 19
   Compositional Differences ......................................................................................... 20
   Effect on Milk Processing and Products ...................................................................... 22
   Effects of Homogenization ......................................................................................... 23
   INFRARED SPECTROSCOPY ....................................................................................... 23
   Challenges of MIR and Milk ....................................................................................... 25
   Genetic Variability of the Spectra ................................................................................ 26
   PREDICTING NOVEL TRAITS USING MIR ................................................................... 26
   Predicting Milk Components ...................................................................................... 28
   Predicting Technological Properties .......................................................................... 30
   Predicting Cow Systemic Traits .................................................................................. 31
   SUMMARY ...................................................................................................................... 32

3. PREDICTION OF MILK FATTY ACID CONTENT WITH MID-INFRARED SPECTROSCOPY USING DIFFERENTLY DISTRIBUTED CALIBRATION SETS ........................................................................................................ 35
   ABSTRACT ...................................................................................................................... 35
   INTRODUCTION ............................................................................................................. 36
   MATERIALS AND METHODS .......................................................................................... 37
   Milk Sampling ............................................................................................................. 37
   Milk Analysis .............................................................................................................. 38
   Data .............................................................................................................................. 39
   MIR Prediction Calibration Sets .................................................................................. 39
   MIR Prediction Models ............................................................................................... 40
   RESULTS AND DISCUSSION ....................................................................................... 41
4. GENETIC CORRELATIONS OF MID-INFRARED PREDICTED MILK FATTY ACID GROUPS WITH MILK PRODUCTION TRAITS ............................................ 51

ABSTRACT .............................................................................................. 51
INTRODUCTION ..................................................................................... 52
MATERIALS AND METHODS ................................................................ 53
  Data ..................................................................................................... 53
  Model .................................................................................................. 54
RESULTS AND DISCUSSION ................................................................. 56
  Phenotypic Description ...................................................................... 56
  Heritabilities ....................................................................................... 56
  Genetic Correlations ......................................................................... 57
  Daily Genetic Correlations .................................................................. 58
CONCLUSIONS .................................................................................... 63

5. VARIATION IN FAT GLOBULE SIZE IN BOVINE MILK AND ITS PREDICTION USING MID-INFRARED SPECTROSCOPY ............................................. 70

ABSTRACT .............................................................................................. 70
INTRODUCTION ..................................................................................... 71
MATERIALS AND METHODS ................................................................ 72
  Data Collection .................................................................................. 72
  Phenotypic Analysis .......................................................................... 73
  Predictive MIR Model ......................................................................... 74
  Correlation with Production Traits ..................................................... 75
RESULTS AND DISCUSSION ................................................................. 75
  Phenotypic Analysis .......................................................................... 75
  MFG Calibration Equations ................................................................. 77
  Correlation with Production Traits ..................................................... 79
CONCLUSIONS .................................................................................... 80

6. GENETIC PARAMETERS OF BOVINE MILK FAT GLOBULE SIZE, MID-INFRARED PREDICTED FAT GLOBULE SIZE, AND FAT AND PROTEIN CONTENTS .......................................................... 89

ABSTRACT .............................................................................................. 89
INTRODUCTION ..................................................................................... 89
MATERIALS AND METHODS ................................................................ 91
  Data ..................................................................................................... 91
  Genetic Analysis ................................................................................. 92
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>RESULTS AND DISCUSSION</td>
<td>93</td>
</tr>
<tr>
<td>Descriptive Statistics</td>
<td>93</td>
</tr>
<tr>
<td>Heritabilities and Repeatabilities</td>
<td>93</td>
</tr>
<tr>
<td>Genetic Correlations</td>
<td>95</td>
</tr>
<tr>
<td>CONCLUSIONS</td>
<td>97</td>
</tr>
<tr>
<td>7. GENERAL DISCUSSION AND CONCLUSIONS</td>
<td>100</td>
</tr>
<tr>
<td>MILK FAT GLOBULE SIZE</td>
<td>100</td>
</tr>
<tr>
<td>MILK FATTY ACIDS</td>
<td>101</td>
</tr>
<tr>
<td>FUTURE OPPORTUNITIES</td>
<td>101</td>
</tr>
<tr>
<td>LIMITATIONS</td>
<td>103</td>
</tr>
<tr>
<td>CONCLUSIONS</td>
<td>103</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>105</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 2.1. Main fatty acids by percent weight of total in triacylglycerols in bovine milk (Christie, 1995) .................................................................................................................. 34

Table 2.2. Gross composition of cow milk fat globule membrane (Keenan and Patton, 1995) ... 34

Table 2.3. Lipids of the milk fat globule membrane (Keenan and Patton, 1995) ............................ 34

Table 3.1. The number of cows, herds, and samples after editing with fatty acid content expressed on a fat (g/100g) and milk (g/dL) basis by breed of cow ............................................. 47

Table 3.2. Descriptive statistics of gas chromatography determined fatty acid content on a fat (g/100g; N = 1,919) and milk (g/dL; N = 1,813) basis .......................................................... 48

Table 3.3. Fitting statistics of each calibration equation estimating fatty acid concentrations using the calibration datasets expressed as g/100g fat (F), g/dL milk (M), ln(g/dL+1) (LN), and the subsets (S) ........................................................................................................ 49

Table 3.4. Measured skewness and excess kurtosis of the distribution of the calibration samples for fatty acids expressed as g/100g fat (F), g/dL milk (M), ln(g/dL+1) (LN), and the subsets (S) ........................................................................................................ 50

Table 4.1. Individual fatty acids included in the defined short-chain (SCFA), medium-chain (MCFA), long-chain (LCFA), saturated (SFA), and unsaturated (UFA) groups of fatty acid traits ........................................................................................................ 64

Table 4.2. Descriptive statistics of test-day records in the edited dataset (N = 49,127) for short-chain (SCFA), medium-chain (MCFA), long-chain (LCFA), saturated (SFA), and unsaturated (UFA) fatty acid groups (ln(g/dL) and milk, fat, and protein yield, fat and protein percentage, fat to protein ratio (F:P) and SCS .......... 64

Table 4.3. Phenotypic correlations for short-chain (SCFA), medium-chain (MCFA), long-chain (LCFA), saturated (SFA), and unsaturated (UFA) fatty acid groups and milk, fats and protein yield, fat and protein percentage, fat to protein ratio (F:P) and SCS ........................................................................................................ 64

Table 4.4. Average daily heritabilities ($h^2$) and posterior standard deviations (SD) of short-chain (SCFA), medium-chain (MCFA), long-chain (LCFA), saturated (SFA), and unsaturated (UFA) fatty acid groups and milk, fats and protein yield, fat and protein percentage, fat to protein ratio (F:P) and SCS ........................................................................................................ 64

Table 4.5. Average daily genetic correlations between short-chain (SCFA), medium-chain (MCFA), long-chain (LCFA), saturated (SFA), and unsaturated (UFA) fatty acid groups and milk, fats and protein yield, fat and protein percentage, fat to protein ratio (F:P) and SCS ........................................................................................................ 65

Table 5.1. Number of cows, records, and herds and the descriptive statistics of measured milk fat globule volume moment mean (D[4,3]) and surface moment mean (D[3,2]) for each breed ........................................................................................................ 82

Table 5.2. Descriptive statistics of measured milk fat globule volume moment mean (D[4,3]) and surface moment mean (D[3,2]) used in prediction models, predicted milk fat
globules volume moment mean (pD[4,3]) and surface moment mean (pD[3,2]),
and milk production traits

Table 5.3. Fitting statistics of prediction models for milk fat globule volume moment mean
(D[4,3]) and surface moment mean (D[3,2])

Table 5.4. Pearson correlation coefficients between measured milk fat globule volume
moment mean (D[4,3]) and surface moment mean (D[3,2]), predicted milk fat
globule volume moment mean (pD[4,3]) and surface moment mean (pD[3,2]),
and milk production traits

Table 6.1. Descriptive statistics of milk fat globule volume moment mean (D[4,3]), surface
moment mean (D[3,2]), predicted volume moment mean (pD[4,3]), predicted
surface moment mean (pD[3,2]), and milk fat and protein percentage

Table 6.2. Estimated heritabilities ($h^2$), their associated standard errors (SE), and
repeatabilities ($r$) of milk fat globule volume moment mean (D[4,3]), surface
moment mean (D[3,2]), predicted volume moment mean (pD[4,3]), predicted
surface moment mean (pD[3,2]), and milk fat and protein percentage

Table 6.3. Estimated additive genetic ($\sigma^2_a$), permanent environment ($\sigma^2_{pe}$), residual ($\sigma^2_e$), and
total phenotypic ($\sigma^2_p$) variances of milk fat globule volume moment mean
(D[4,3]), surface moment mean (D[3,2]), predicted volume moment mean
(pD[4,3]), predicted surface moment mean (pD[3,2]), and milk fat and protein
percentage

Table 6.4. Genetic correlations with standard errors in parentheses of milk fat globule
volume moment mean (D[4,3]), surface moment mean (D[3,2]), predicted volume
moment mean (pD[4,3]), predicted surface moment mean (pD[3,2]), and milk fat
and protein percentage
LIST OF FIGURES

Figure 4.1. Daily genetic correlations for short-chain (SCFA), medium-chain (MCFA), long-chain (LCFA), saturated (SFA), and unsaturated (UFA) fatty acid groups averaged across the three analyses, and fat percentage (FAT %) with milk yield .... 66

Figure 4.2. Daily genetic correlations for short-chain (SCFA), medium-chain (MCFA), long-chain (LCFA), saturated (SFA), and unsaturated (UFA) fatty acid groups with fat yield ................................................................. 66

Figure 4.3. Daily genetic correlations for short-chain (SCFA), medium-chain (MCFA), long-chain (LCFA), saturated (SFA), and unsaturated (UFA) fatty acid groups with fat percentage ................................................. 67

Figure 4.4. Daily genetic correlations for short-chain (SCFA), medium-chain (MCFA), long-chain (LCFA), saturated (SFA), and unsaturated (UFA) fatty acid groups with protein yield .................................................................... 67

Figure 4.5. Daily genetic correlations for short-chain (SCFA), medium-chain (MCFA), long-chain (LCFA), saturated (SFA), and unsaturated (UFA) fatty acid groups and fat percentage (FAT %) with protein percentage ........................................ 68

Figure 4.6. Daily genetic correlations for short-chain (SCFA), medium-chain (MCFA), long-chain (LCFA), saturated (SFA), and unsaturated (UFA) fatty acid groups with fat to protein ratio .................................................................... 68

Figure 4.7. Daily genetic correlations for short-chain (SCFA), medium-chain (MCFA), long-chain (LCFA), saturated (SFA), and unsaturated (UFA) fatty acid groups averaged across the three analyses and fat percentage (FAT %) with somatic cell score ............................................................. 69

Figure 5.1. Least squares means of milk fat globule size measured as (a) volume moment mean (D[4,3]), and (b) surface moment mean (D[3,2]) for season ................................................. 84

Figure 5.2. Least squares means of milk fat globule size measured as (a) volume moment mean (D[4,3]), and (b) surface moment mean (D[3,2]) for Holstein (HO; N = 1,406 samples), Ayrshire (AY; N = 171), Brown Swiss (BS; N = 60), and Jersey (JE; N = 189) breeds ........................................................................ 85

Figure 5.3. Least squares means of milk fat globule size measured as volume moment mean (D[4,3]), and surface moment mean (D[3,2]) for classes of days in milk ....................... 86

Figure 5.4. Predicted vs. measured milk fat globule size measured as (a) volume moment mean (D[4,3]), and (b) surface moment mean (D[3,2])................................................................. 87

Figure 5.5. Predicted vs. measured milk fat globule size measured as surface moment mean (D[3,2]) for samples with a common (4.0%) milk fat content.............................................. 88
Chapter 1
GENERAL INTRODUCTION

INTRODUCTION

Milk has evolved to be a complete food to provide nourishment, vigour, and protection to the suckling mammalian neonate. Cow milk and its derived products have also become an important source of nutrition in many human populations. Consumers are more often basing their food choices on their health aspects, wanting food products that not only provide nutrients, but also encourage good overall health or prevent disease. High milk quality is a priority for the dairy industry and its quality can be considered from nutritional, sensory, hygienic-toxicological, and technological points of view. All of these features contribute to producing a wholesome, enjoyable, and most importantly safe product. These properties are mainly determined by the milk composition. Furthermore, there is an economic value to the milk quality and milk composition. There is a great interest in the composition of milk because of its effects on human health through its consumption, as well as its use as an indicator of cow physiological status. It is therefore becoming increasingly necessary for producers to take milk composition into consideration for herd management and in order to adapt their production for consumer and industry needs.

Fat is a particularly key component in milk due to its role in the economics, nutrition, flavour, and physico-chemical properties of milk and milk products. Two important characteristics of milk fat with nutritional, technological, and animal implications are the fatty acid (FA) composition and the size of the milk fat globules (MFG). Milk fat is high in saturated FA, which is a concern to consumers due to the known link between saturated fat intake and increased risk of cardiovascular disease (Haug, 2007). There are also many milk FAs that favourably influence health and may have anticarcinogenic, antiallergic, antimicrobial, and anti-inflammatory effects (Williams, 2000; Parodi, 2004; Haug, 2007). There may be an opportunity to change and optimize the milk FA profile to better its composition for human health. Changes to the FA contents can also alter the technological properties of milk, thereby affecting products of milk processing (Chen et al., 2004). Furthermore, since milk FAs are being produced directly by the cow, as well as being sourced from dietary and adipose lipids, changes in milk FA could be used as an indicator for changes in the physiology of the cow.
The membrane of MFG contains many health beneficial components (Spitsberg, 2000). The milk of cows producing on average smaller MFG, for a given fat content, contains more of this membrane material. The size of the MFG is also an important characteristic for dairy processing and can impact products such as cheese (Michalski et al., 2003). Therefore, it would be of interest to examine the phenotypic and genetic variation of FA composition and MFG size in the Canadian dairy population in order to further assess their relationships with other factors and traits and assess if selection could be performed. Unfortunately, these traits are expensive and difficult to obtain large numbers of phenotypes for, especially on a routine basis, which would be required for genetic selection.

Nowadays, milk composition testing is performed by certified milk recording laboratories for payment, quality control, herd management, and genetic selection purposes. Mid-infrared (MIR) spectrometry is routinely used to simultaneously provide information on a variety of compositional parameters determined from the obtained spectra of the milk sample in a rapid and inexpensive manner. Fat and protein percentages, milk-urea-nitrogen (MUN), and beta-hydroxybuterate (BHB) are determined routinely from MIR spectra and reported for bulk farm milk samples and samples from individual cows for herd management and genetic evaluation. In Canada, official milk recording services within Dairy Herd Improvement (DHI) are completed by Valacta (Sainte-Anne-de-Bellevue, Québec) for Québec and Atlantic Canada, and CanWest DHI (Guelph, Ontario) for Ontario and Western Canada. In 2015, 409,742 cows (43% of Canadian dairy cows) across Canada were enrolled in official supervised milk recording programs (Canadian Dairy Information Centre, Agriculture and Agri-Food Canada). A minimum of 8 or 10 tests is required per year for official milk recording. An additional 284,523 cows (30% of Canadian dairy cows) were enrolled in management milk recording services, which are unsupervised and require no minimum number of tests per year. Dairy herd improvement milk recording programs help to develop an extensive database used for genetic evaluation.

Traditionally, selection in the dairy cattle industry has focused largely on production traits. More recently, broader, more balanced selection goals have taken health and reproduction traits into consideration to complement the selection on production traits (Miglior et al., 2005). Increased public interest in milk quality and milk production practices may require new traits to be defined and examined, and selection objectives to progress further. The interest in several new traits is already present, but there is currently no system in use to regularly collect these
phenotypes on a large enough scale. This lack of data collection may be due to the current unavailability of technology or the under-utilization of the available technology. The spectra generated during milk recording for the vast number of milk samples possess an immense amount of information in regards to the milk sample, and thus the cow, beyond the traits already determined. Therefore, many researchers have recently began to look at using MIR spectroscopy to phenotype additional, more detailed milk composition, milk properties, and cow characteristics (De Marchi et al., 2014). These traits include detailed milk fat and protein contents (Soyeurt et al., 2006a; McDermott et al., 2016), milk coagulation properties (De Marchi et al., 2013), and cow energy status and efficiency (McParland et al., 2011, 2014). MIR technology provides an opportunity to obtain these phenotypes for large numbers of milk samples and cows at a low cost with technology already used in regular milk recording. Mid-infrared spectra are already generated within DHI milk recording for all milk samples and this data can be used in innovative ways through the prediction of additional novel traits. Several countries, including Canada, have begun to retain the spectra of milk samples to analyze post hoc as new traits derived from MIR spectra are constructed. This would allow for predicted traits on historical as well as incoming spectra.

The development of prediction equations for milk FA composition and MFG size using MIR spectroscopy for use in the Canadian dairy industry would be valuable. Beyond assessing the accuracy of predictions for these traits from MIR spectra, understanding the relationship between the predictions and other known traits, such as production and health traits, is imperative before they can be used. This includes understanding both the phenotypic and genetic variation that exists, and how their alteration could influence other milk components, milk production, and the cow.

OBJECTIVES

The objectives of this thesis were to examine the phenotypic and genetic variability of milk fatty acids and average milk fat globule size and to assess the applicability of MIR spectral predictions of these traits for selection purposes.

Chapter 2 gives an analysis of the current literature related to milk fatty acids, milk fat globule size, and the use of mid-infrared spectroscopy technology as it applies to milk. Chapter 3 looks to develop prediction equations for milk fatty acid contents using a calibration set of gas
chromatography determined fatty acid profiles and mid-infrared spectral data. Modifications to the calibration set in terms of how the fatty acids are expressed and which samples are included are tested in order to improve the calibration. In Chapter 4, the fatty acid prediction equations developed in Chapter 3 are used on a subset of saved mid-infrared spectra from first-parity Holstein cows to obtain predicted contents of groups of fatty acids. Multiple trait random regression test-day models are used to estimate the genetic correlation between the groups of predicted fatty acids and milk production traits.

Chapter 5 investigates the phenotypic variation observed in average milk fat globule size. The effects of herd, breed, days in milk, season, milking period, age at calving, parity, and individual animal on milk fat globule size are examined. As well, the ability to predict average milk fat globule size from mid-infrared spectra is studied. Production trait records are obtained and phenotypic correlations among the measured and predicted average milk fat globule size and production traits are estimated. In Chapter 6, a comparison of the genetic variation of the gold standard laboratory measured milk fat globule size and the MIR predicted milk fat globule size is performed. The heritability is estimated for both measured and predicted MFG size along with the genetic correlation between them and with production traits. Chapter 7 is a general discussion of the results presented in the previous chapters and the future opportunities of the research and its limitations are reported.
Chapter 2
LITERATURE REVIEW

MILK LIPIDS

Bovine milk is comprised of approximately 87% water, 4.6% lactose, 3.4% protein, 4.2% fat, 0.8% minerals, and 0.1% vitamins (Månsson, 2008). The lipid content in cow milk is quite variable and typically ranges from 33 to 47 g/L (Christie, 1995). The milk fat component is comprised primarily of triacylglycerols, representing about 98% of total lipids, with diacylglycerols, monoacylglycerols, phospholipids, cholesterol, and free fatty acids making up the remaining milk lipids (Jensen, 1995). Triacylglycerols are comprised of three individual fatty acids (FA), a carboxylic acid with a long aliphatic tail, and a glycerol backbone. The attachment of FA to the three positions of the triacylglycerol is not random, and FA tend to preferentially attach at certain positions (Jensen, 2002). This limits the variation observed in triacylglycerol structures, and only about 22 triacylglycerol structures are found at >1 mol % in milk fat (Jensen, 2002). The different FA incorporated within the triacylglycerol can vary in chain length, saturation, and arrangement depending on their availability, which has a major influence on the properties of the milk fat.

MILK FATTY ACIDS

Over 400 different FA have been identified in milk fat (Jensen et al., 1991). However, most of these FA appear in trace amounts, and there are only around 12 FA present at or above 1% concentration (Jensen, 2002). The main FA appearing in milk by percent weight are shown in Table 2.1. Approximately 70% of FA are saturated and 30% are unsaturated.

Origins of Milk Fatty Acids

The FA in milk fat originate from two sources: from de novo synthesis within the mammary glands and from the uptake of preformed plasma lipids by the mammary gland. Plasma lipids are released from body fat stores, formed in the rumen from biohydrogenation or bacterial degradation, or arise directly from the diet. Short and medium-chain length FA (containing 4 to 14 carbon atoms) are almost exclusively synthesized de novo in the mammary gland. About half of palmitic acid (C16:0) is also synthesized de novo. Short- and medium-chain
FA are synthesized in mammary epithelial cells using acetate and BHB generated by the microbial fermentation of cellulose and hemicellulose in the rumen (Cozma et al., 2013). The two key enzymes in FA synthesis are acetyl-CoA carboxylase (ACC) and fatty acid synthetase (FAS). Acetate and BHB contribute equally to the initial 4-carbon atoms. Acetate is transformed into acetyl-CoA to be used in chain length extension and BHB is changed to butyryl-CoA and then incorporated (Cozma et al., 2013).

Long-chain FA (16 and 18 carbon atoms) are provided to the mammary gland from the blood stream and originate from lipoproteins and non-esterified FA that originate from the dietary and microbial lipids absorbed in the digestive tract and from lipolysis of adipose lipids (Cozma et al., 2013). Palmitic (C16:0) and stearic (C18:0) acids pass through the rumen unchanged, while biohydrogenation of unsaturated FA by rumen microorganisms result primarily in stearic acid and some oleic acid (C18:1). In the mammary epithelial cells, the enzyme stearoyl-CoA desaturase, catalyses the oxidation of fatty acyl CoA esters, causing a cis-double bond between carbons 9 and 10 for FA with a chain length from 10 to 18 carbons (Jensen, 2002). Odd-chain FA such as pentadecanoic acid (15:0) and heptadecanoic acid (C17:0) are synthesized by the rumen bacterial flora.

**Potential Health Effects**

Milk fat is an important contributor of nutrition in the diet of many human populations. The beneficial and negative health effects of milk fat components have been discussed extensively (e.g., Williams, 2000; Jensen, 2002; Parodi, 2004; Terpstra, 2004; Haug, 2007; Shingfield, 2008; German, 2009).

There is a general negative opinion of milk fat amongst consumers largely due to its high proportion of saturated FA, which have been associated with increased risk of atherosclerosis, cardiovascular disease, obesity, and weight gain (Haug, 2007). The consumption of saturated FA increases total and low-density lipoprotein (LDL) cholesterol in the blood. Largely, epidemiological studies have studied the effects of fatty acid intake using broad fatty acid groupings. However, the effects of the individual FA within these groups may not be the same. Intake of the saturated FA C14:0 and C16:0, have been shown to increase LDL cholesterol concentrations in humans (Temme et al., 1996), which is associated with cardiovascular disease. Contrastingly, the other major saturated FA, C18:0, and very short-chain FA show neutral effects
on LDL cholesterol levels (Grundy, 1994; Hu et al., 1999). The adverse effects on blood LDL levels may be mitigated by other milk fats such as HDL-increasing saturated FA and polyunsaturated FA (Hu et al., 1999).

Trans-FA have also been linked to increased risk of coronary heart disease. The effects of trans-FA may not however be the same for all trans-FA and may be isomer-dependent. Ruminal derived trans-FA may not have the same impacts as those created from the partial hydrogenation of vegetable oils (Shingfield, 2008). The relationship between coronary heart disease risk and trans-FA from ruminants was not observed by Jakobsen et al. (2008).

Of particular interest, due to its potential health benefits, is conjugated linoleic acid (CLA). There is evidence that CLA may have anticarcinogenic (Parodi, 1999), body fat-lowering (Park et al., 1997), and artherosclerosis-reducing properties (Lee et al., 1994). In mice, Ip et al. (1991) found CLA given prior to carcinogen treatment modulated tumour growth. The long-chain omega-3 polyunsaturated FA and its ratio with omega-6 polyunsaturated FA may be of interest to increase as well. High intakes of omega-6 FA may be related to breast, colon, and prostate cancer, while omega-3 FA have protective properties (Bartsch et al., 1999).

The literature on the effects of milk fat components on human health contains many contradicting studies for the same components as well as debatable results. Diet is only one contributor to human health and disease, and the effects of other factors such as age, social class, gender, and lifestyle choices (smoking, exercise, etc.) may not be fully controlled or adjusted for in human studies. Many studies also look at the effects of milk fat components using non-human models and the same effects may not be observed for humans or for normally consumed quantities of the components (Terpstra, 2004).

Current recommendations are for the public to decrease their consumption of saturated fats and consume more unsaturated FA (FAO, 2010). A reduction in the proportion of saturated FA in bovine milk may therefore be of great interest. More evidence of the effects of individual FA is required before ideal FA contents can be determined.

Different milk FA may also have benefits for the udder health of dairy cows. Hogan et al. (1987) found the free FA C12:0, C14:0, and 18-carbon FA inhibited the growth of various gram-positive, mastitis producing pathogens in vitro. These bacteriostatic properties were dose dependent and it is unclear if the benefits would be observed at an animal level.
Technological and Sensory Properties

The chain length and saturation level of FA in milk can have dramatic consequences for technological and sensory properties of milk fat. Largely affected are the melting temperature and crystallization of milk fat, which can have an impact on many milk products. In general, melting point increases as the molecular weight increases. Therefore, long-chain FA will have higher melting points than short-chain FA, as illustrated by the fact that C16:0 has a melting temperature 6.5 degrees lower than that of C18:0 (Berg et al., 2002). As a group, unsaturated FA have lower melting points than saturated FA due to the weaker intermolecular interaction of unsaturated molecules. C18:0 melts at 69.6°C while 18:1c9 melts at 13.4°C (Berg et al., 2002). Relatedly, the FA included in triacylglyceride structures affects crystallization behaviour, which impacts rheological and plasticity properties of milk products.

Butter with a softer texture and improved spreadability may be created from milk fat with increased unsaturated and short-chain FA (Bobe et al., 2007). Chen et al. (2004) found different measures of texture for butter, ice cream, and Provolone cheese were altered by the level of unsaturated FA in the milk used to produce them. There is a risk that with too high of levels of polyunsaturated FA in milk products, the fat may be too fluid and “oil off” at higher temperatures (Wood et al., 1975). Milk products high in polyunsaturated FA also may have reduced oxidative stability, which can produce off-flavours (Wong et al., 1973).

Factors Influencing Fatty Acid Composition

Diet. Cow nutrition is the primary factor affecting FA composition of milk and the topic has been reviewed extensively (e.g., Palmquist et al., 1993; Chilliard et al., 2000; Vlaeminck et al., 2006). Thus, altering the diet of a lactating cow is a popular and effective means for changing the FA content of her milk. The feed factors that can alter milk FA contents include season and region of feed production, feeding various fats and oils, supplemented oils, and protected oils (Jensen, 2002).

Lower proportions of saturated FA and higher proportions of unsaturated FA in milk can be achieved by replacing forage and concentrate ingredients with plant oils, oilseeds, or lipid supplements that are resistant to metabolism in the rumen (Kliem and Shingfield, 2016). In the rumen, there is extensive lipolysis and biohydrogenation of dietary unsaturated FA, which explains the low efficiency of the transfer of unsaturated FA from the diet into milk. Transfer
efficiency can depend on the specific FA, dietary source, basal diet, rumen environment, and if the FA are used elsewhere by the cow and thus not made available to the mammary gland (Lock and Bauman, 2004; Woods and Fearon, 2009; Kliem and Shingfield, 2016). The degradation of unsaturated FA in the rumen can be reduced by protecting dietary lipids: by (i) encapsulating them in formaldehyde proteins (Ashes et al., 1979); or (ii) using the calcium salts of FA (Ferlay et al., 1993).

The amount of CLA in milk can be increased through supplements with plant oils and oilseeds enriched in 18:2 n-6 and 18:3 n-3 (Kliem and Shingfield, 2016). Milk cis-9, trans-11 CLA increases are due to the accumulation of trans-11 C18:1 in the rumen, and its availability for desaturation in the mammary gland (Palmquist et al., 2005). One of the most effective ways of doing this is through the feeding of fish oil (Chilliard, 2000). Donovan et al. (2000) fed a menhaden fish oil diet at 0, 1, 2, and 3% of ration to lactating Holstein cows and found a 2% fish oil diet resulted in the highest concentrations of cis-9, trans-11 CLA and trans-11 C18:1 in milk. Notably, high quantities of fish oil and marine algae supplements may cause a decrease in milk fat synthesis by altering rumen biohydrogenation (Donovan et al., 2000; Lock and Bauman, 2004). Grass is a main source of C18:3 for cows. The amount of fat and the fat composition of grass vary according to its maturity (Chilliard, 2000). Kelly et al. (1998) found Holsteins grazing on fresh pasture produced a significantly greater proportion of CLA in their milk than a control group consuming a total mixed diet (1.09% vs. 0.46% CLA of fat weight). However, milk, fat, and protein yields were also significantly lower for the grazing group.

**Animal Factors.** The FA composition of milk varies between cattle breeds (Arnould and Soyeurt, 2009). Soyeurt et al. (2006b) examined the FA quantities in milk of different breeds and found breed did have an effect on FA composition, but many of these differences are likely more related to differentiations in total fat content between the breeds. When expressed in g/100g of fat, Montbeliarde did have greater proportions of C12:0 and C14:0 compared to Holstein, but no significant difference was observed between other examined breeds. Using samples from a single farm, DePeters et al. (1995) discovered that Jersey cows tended to have higher proportions of C6:0, and C8:0 than Holsteins; higher proportions of C10:0, C12:0, and C18:0; and lower proportions of C18:1c9,10 than both Holsteins and Brown Swiss cows. Most differences in FA composition reported in the literature are between Holstein and Jersey breeds for short-chain FA
The composition of milk fat is also affected by the genetics of the individual animal, as reviewed by Arnould and Soyeurt (2009). The heritabilities of the FA tend to reflect their origin and biological pathway. Short- and medium-chain FA synthesized \textit{de novo} in the mammary tissue are more heritable than long-chain FA (Bobe et al., 2008; Stoop et al., 2008; Garnsworthy et al., 2010; Bastin et al., 2011). However, Mele et al. (2009) found low and similar heritability estimates of 0.07, 0.03, and 0.08 for C14:0, C16:0, and C18:0, respectively. The heritabilities of the desaturated FA were slightly higher. Soyeurt et al. (2008) estimated average daily heritabilities of 0.42 and 0.14 for saturated and monounsaturated FA expressed in g/100g of milk, respectively, and 0.24 and 0.27 for the same FA groups when expressed in g/100g of fat. Bobe et al. (2008) found, in general, greater heritabilities for individual FA and FA grouped by saturation when they were expressed in g/L of milk compared to a percent weight of total FA.

Stoop et al. (2008) reported mostly strong, positive genetic correlations among C6:0 to C14:0 FA, noting they are synthesized by the same pathway. The genetic correlations between FA with a chain length lower than 16 with 18-carbon FA are weaker as a result of their distinctive origins. Bastin et al. (2011) approximated the genetic correlations between individual and groups of FA (g/dL of milk) using correlations among daily breeding values for the traits. They also noted strong genetic correlations between FA of similar origins, with genetic correlations ranging from 0.64 to 0.99 among \textit{de novo} synthesized FA and ranging from 0.27 to 0.66 between these FA and the 18-carbon length FA. The genetic correlation for saturated FA with short-, medium-, and long-chain FA were 0.92, 0.99, and 0.66, respectively, showing that most short- and medium-chain FA are also in the saturated FA grouping.

Stoop et al. (2008) found that milk fat percentage was negatively correlated with unsaturated 18-carbon FA, suggesting selection for increased fat content will decrease these FA. Similar results were found by Mele et al. (2009) for fat percentage with C18:1\textit{t}11 and CLA (correlation coefficients of 0.69 and 0.55, respectively), but not between fat percentage and C18:1\textit{c}11 (correlation coefficient of 0.02). Soyeurt et al. (2008) found positive daily genetic correlations (ranging from 0.23 to 0.62) between fat percentage and saturated FA, while, apart from the initial stage of lactation, negative daily genetic correlations were observed between fat percentage and monounsaturated FA. Based on genetic correlations estimated using the
correlation of daily estimated breeding values between FA and production traits, Bastin et al. (2011) found that the analyzed FA were negatively correlated with milk yield and protein yield, but positively correlated with fat yield, fat percentage, and protein percentage. Penasa et al. (2015) reported very similar genetic correlations for FA grouped by saturation amount. Bastin et al. (2011) noted correlations with milk yield varied across days in milk, with many FA having genetic correlations with milk yield closer to zero at the start of lactation and becoming more negative thereafter. These changes present the awareness that selecting for milk yield at various time points in a lactation could have different effects on the FA content.

Differences that are observed in the estimated genetic parameters are attributable to differences in data sets and methodologies. There is a range in sample sizes, FA measuring methodologies (gas chromatography, MIR spectroscopy), measurement units (per gram of fat, per volume of milk), and in the genetic models used affecting the results reported (Arnould and Soyeurt, 2009).

Genomic studies have also been performed to identify variants in the genome associated with FA contents. The DGAT1 and SCD1 genes are known to have major roles in milk fat composition. The DGAT1 enzyme is involved in the last step of the synthesis of triacylglycerol and has a major effect on fat content (Grisart et al., 2002). As well, variants in DGAT1 have been shown to be connected with more saturated FA and C16:0, and smaller fractions of C14:0, unsaturated C18, and CLA (Schennink et al., 2007). The SCD1 enzyme catalyzes the desaturation of several saturated FA into unsaturated forms.

**Fatty Acid Profile as an Indicator of Other Traits**

The FA profile of a cow’s milk may provide an indication of the physiological state of that cow at a particular time point. Of particular interest is using the FA profile as a connection to energy supply and the metabolic status of the cow. The energy balance of the cow is the difference between the net energy intake and the net energy output needed for maintenance and milk production. Fatty acid composition is thus strongly dependent on the stage of lactation. At the beginning of lactation, cows are in negative energy balance causing an increase in the mobilization of body reserves. As a result, long-chain FA coming from the adipose tissues will be incorporated into the milk fat (Palmquist et al., 1993). As well, the de novo synthesis of short-chain FA in the mammary gland can be inhibited by the high uptake of the long-chain FA,
decreasing the short-chain FA contents in the milk fat (Palmquist et al., 1993). Stoop et al. (2009) found a relationship between negative energy balance and an increase in C16:0 and C18:0 in milk, which may be a result of the mobilization of body fat reserves. They also noted a decrease in odd-chain FA (C5:0 to C15:0), due to reduced availability of glycogenic (C3) products required for milk production or the inhibition of the synthesis of these FA. When the energy balance improved as lactation progressed, Gross et al. (2011) observed an increase in short- and medium-chain FA, and a decline in long-chain FA, 18:1c9 especially.

The energy balance of the cow has been associated with cow health (Ingvartsen et al., 2003). Ketosis and fatty liver may be considered as the failure of the cow to adapt to negative energy balance (Herdt, 2000). Detection of FA profile changes resulting from the mobilization of adipose may aid in ketosis diagnosis (Arnould et al., 2013). Van Haelst et al. (2008) studied cows with subclinical ketosis diagnosed via BHB concentrations, and noted significant decreases in long-chain FA and C18:1c8 from pre-diagnosis to diagnosis and from diagnosis to post-diagnosis time points. These observed differences, however, are very likely due to the effect of lactation stage and not to the effect of ketosis. Additionally, in early lactation, Collard et al. (2000) reported significant phenotypic correlations between mean energy balance and laminitis occurrences. They also found significant correlations for minimum energy balance with locomotive problems, laminitis, and digestive order occurrences.

Negative energy balance may also inhibit the reproductive performance of dairy cows (Jorritsma et al., 2003; Pryce et al., 2004). Energy balance is related to time to first ovulation (Butler and Smith, 1989; Beam and Butler, 1999) and lower conception rates (Villa-Godoy et al., 1988). Using Czech Fleckvieh cows from a single herd, Stádník et al. (2015) found significant differences for both number of services per conception and days open between cows producing either high or low quantities of saturated FA, as well as monounsaturated FA, during the first 5 weeks of lactation. No correlation was found between the FA groups in milk and calving to first service interval. Bastin et al. (2012) looked at the genetic correlations between the major FA in milk and days open and acknowledged the pattern observed in the genetic correlations across the lactation are likely related to the changing physiological state of the cow. They reported a positive genetic correlation in early lactation but negative genetic correlation after 100 days in milk for days open with unsaturated FA, monounsaturated FA, long-chain FA, C18:0, and C18:1c9. For the de novo synthesized FA, C6:0 to C14:0, the genetic correlation with days open
was more negative in the beginning stage of lactation than after 100 days. These results suggest an opportunity to select for fertility traits with low heritabilities using FA as indicator traits. It will also be important to consider these genetic correlations to avoid negative correlated responses if direct selection for FA composition were to occur.

Randolph and Erwin (1974) examined quarter milk samples of 10 dairy cows for mastitis. In the positive milk, there were higher concentrations of the short-chain FA and lower concentrations of C16:0 and C18:0 per gram of fat than was found in the negative milk. Conversely, Massart-Leën et al. (1994) did not note any changes in triacylglycerol FA composition between healthy milk quarters and induced Escherichia coli mastitis milk quarters. The most marked difference for FA in mastitic milk is an increase in free FA concentrations produced by the lipolysis of milk fat (Gudding, 1982), rather than a change in the FA profile.

Fatty acid profiles may also be used as an indicator for the methane production of lactating cows (Chilliard et al., 2009; Dijkstra et al., 2011). The relationship between milk FA composition and methane production may be due to the common biochemical pathways of methane, acetate, and butyrate in the rumen, or a result of dietary fats, through the conversion of pyruvate into acetate, butyrate, propionate, carbon dioxide, and hydrogen, which is later used to produce methane (Chilliard et al., 2009). Using 50 observations of Holstein-Friesian cows fed different diets, Dijkstra et al. (2011) found that methane production was positively correlated with milk C8:0, C10:0, C11:0, C14:0 iso, C15:0 iso, C16:0, and C17:0 anteiso contents, and negatively correlated with milk C17:0 iso, C17:1c9, C18:1c9, C18:1t10,t11, C18:1c11, C18:1c12 and C18:1c4,t14. They were able to predict methane production in g/kg of dry matter from individual milk FA with an $R^2$ of 0.73. Chilliard et al. (2009) also found methane production to be positively correlated with milk C6:0 to C16:0. Negative correlations were found for several unsaturated 18-carbon FA, which are biohydrogenation intermediates. Using a multiple regression model which included C16:0, C18:1c14, C14:1c9, and C18:2n-6 milk FA contents along with forage intake, Chilliard et al. (2009) was able to accurately predict methane output in g/d ($R^2 = 0.953$). However, note that equations using milk FA to predict methane production need to be validated using more animals and populations. Lassen et al. (2016) estimated the genetic correlations between milk FA profiles and methane related traits using 339 Holstein cows with BovineSNP50 genotypes using pedigree and SNP-based relationship matrices. Standard errors of the estimated genetic correlations were very large (mean SE ranged from 0.43
to 0.66), but significant genetic correlations were found between methane related traits and some FA, mostly C13:0, C15:0, and C17:0.

**MILK FAT GLOBULES**

Over 95% of the total milk fat is present in the form of milk fat globules (MFG) (Keenan and Dylewski, 1995). Milk fat is secreted from mammary epithelial cells as spherical MFG composed primarily of a globule of triacylglycerol surrounded by a cellular membrane, called the milk fat globule membrane (MFGM). The MFGM compartmentalizes the triacylglycerol-rich core and stabilizes the globules making them compatible with an aqueous environment. In addition, the MFGM protects the lipid core from lipolysis. There are typically about $10^{10}$ milk fat globules per mL of milk (Walstra et al., 1999).

Understanding the formation and secretion of MFG have been the focus of much research and the subject of many reviews (Mather and Keenan, 1998; Keenan, 2001; Ollivier-Bousquet, 2002; Heid and Keenan, 2005; Argov et al., 2008). Milk fat globule triacylglycerols are synthesized at the rough endoplasmic reticulum, accumulate and released as microlipid droplets into the cytosol, with a surface coat of polar lipids and proteins derived from the endoplasmic reticulum (Mather and Keenan, 1998). The droplets grow in size through coalescence with each other as they migrate from their sites of origin, mostly in the basal and lateral cell regions, to the apical region where growth is most pronounced (Stemberger and Patton, 1981). Fusion between droplets was shown by Valivullah et al. (1988) with evidence of microlipid droplets fusing with each other and with larger lipid droplets. Large lipid droplets have not been shown to fuse together. The mechanisms involved in the fusion of lipid droplets and whether these processes are regulated or random remains speculative. Lipid droplets reach the apical surface of the cell and are gradually coated with plasma membrane. When the droplet is completely enveloped in the apical membrane, it is pinched off and the MFG is expelled into the alveolar lumen.

The materials of the MFGM originate from the endoplasmic reticulum, specialized regions of the apical plasma membrane, and possibly from other intracellular components of mammary epithelial cells. The MFGM has a tripartite structure consisting of an inner surface-active monolayer and a true outer bilayer membrane with a dense protein-rich coat on the inner face of the bilayer membrane (Keenan and Mather, 2002). The thickness of the MFGM varies between 10 and 20 nm (Walstra et al., 1999) and accounts for an estimated 2-6% of the total
mass of the MFG (Singh, 2006). Materials found in the inner coat of the MFGM originate from the endoplasmic reticulum during the initial intracellular formation of the droplet. This monolayer of proteins and polar lipids covers the triacylglycerol-rich core of the globule preventing their coalescence in the cytoplasm. The coating also plays a role in droplet fusion and possibly droplet interaction with the plasma membrane (Keenan and Patton, 1995). The outer bilayer provides the backbone of the MFGM is most likely derived in part, or possibly entirely, from apical plasma membrane. The bilayer has peripheral membrane proteins partially embedded or loosely attached and trans-membrane proteins extending through it (El-Loly et al., 2011).

The complete MFGM includes proteins, phospholipids, glycoproteins, and enzymes distributed asymmetrically within the membrane. The gross composition of the MFGM is shown in Table 2.2. Proteins represent 25-60% of the weight of the MFGM and are very diverse. The MFGM proteins account for 1-2% of the total protein in milk (Riccio, 2004). The major protein component of the bovine MFGM is butyrophilin that accounts for approximately 40% of the total membrane protein in Holstein cow milk (Huppertz et al., 2009) and is only evident in mammary tissue during lactation.

The main classes of lipids in the MFGM are summarized in Table 2.3. The MFGM is largely composed of triacylglycerols and phospholipids in an approximate 2:1 ratio. The fatty acid composition of MFGM triacylglycerols differs from that of the triacylglycerols of the core in that MFGM-associated triacylglycerols contain higher proportions of long-chain saturated fatty acids (Kitchen, 1977). Approximately 60% of the total phospholipids in milk are present in MFG, and are mainly, if not exclusively, associated with the MFGM (Keenan and Patton, 1995). The three main MFGM phospholipids are sphingomyelin, phosphatidylcholine, and phosphatidylethanolamine.

**Potential Health Effects of the Milk Fat Globule Membrane**

The MFGM is valuable to the human diet as it is naturally rich in important minor lipids and glycoproteins. Bovine MFGM has been suggested as a prospective nutraceutical due to its many health-beneficial components (Spitsberg, 2005; Hintze et al., 2011). At a given amount of fat, cows producing small MFG will be producing higher amounts of the beneficial MFGM material. The health benefits of the MFGM can be attained by consuming it alone, as a nutraceutical, in dairy products, or in enriched food products. An extensive examination of the
nutritional properties of the MFGM can be found in reviews by Spitsberg (2005), Dewettinck et al. (2008), El-Loly (2011), and Hintze et al. (2011). Much of the existing research is still speculative and disparities exist. Some MFGM components have been suggested to be harmful to human health. More research in human subjects is necessary to draw more definitive conclusions but studies providing promising results are accumulating.

**Fat Fraction.** The consumption of MFGM may have many health benefits due to the presence of phospholipids. Phospholipids affect various cell functions including growth and development, molecular transport systems, absorption processes, memory, stress responses, development of Alzheimer’s disease, and myelination in the central nervous system (Spitsberg, 2005). The MFGM phospholipids also contain relatively larger quantities of polyunsaturated fatty acids than the triacylglycerols. Polyunsaturated fatty acids have many valuable properties for human health related to cognitive function, growth and development, inflammatory conditions, and blood lipids.

The MFGM is a significant dietary source of sphingolipids. Interest in sphingolipids has increased because of their importance in regulating biological processes like cell growth, differentiation, and senescence. Sphingomyelin, a sphingophospholipid, is the predominant sphingolipid membrane component and it is one of the few membrane phospholipids not synthesized from glycerol. As reviewed by Schmelz (2004) and by Duan (2005), there is a body of evidence indicating the anticarcinogenic mechanisms of sphingolipid metabolism in the colon mainly through the various physiological functions of ceramide and sphingosine, such as inducing cell growth arrest, differentiation, and apoptosis. Although human clinical trials have not been presented, sphingosine and ceramide were able to induce apoptosis in a human adenocarcinoma cell line (Schmelz et al., 1998), suggesting the anticarcinogenic effect may translate to reduced human colon cancer risk. Sphingolipids are also involved in the intestinal uptake of cholesterol and have been shown to have anticholesterolemic effects in animal models. Noh and Koo (2004) reported milk sphingomyelin inhibited the absorption of cholesterol in rats, and did so more effectively than egg sphingomyelin, possibly due to the highly saturated long chains of milk sphingomyelin fatty acyl groups. Conway et al. (2013), using a double-blinded randomized crossover placebo controlled studied in humans, observed a reduction in serum cholesterol, LDL-cholesterol, and triacylglycerol with the consumption of buttermilk.
Phosphatidylserine is a major type of phospholipid present in the MFGM, although present in only small amounts in milk. Phosphatidylserine can attenuate many neuronal effects of aging and it has been shown to have positive effects on memory performance (El-Loly, 2011). Findings on the effectiveness of phosphatidylserine are limited though, and the low concentration in milk gives limited significance to phosphatidylserine ingestion from milk. Research has also suggested phosphatidylserine as having possible ergogenic properties by reducing accumulated oxygen deficit and improving exercise tolerance (Kingsley et al. 2005, 2006). However, again, the levels of supplementation were much greater than that attainable through milk consumption.

Finally, it is believed that phosphatidylcholine can clinically support liver recovery from toxic chemical attack or acute or chronic viral damage (Kidd, 2002). Phosphatidylcholine plays an important role in liver cell activation, proliferation, maturation, and regeneration following damage by supplying components required for key metabolic pathways (Kidd, 2002). Phosphatidylcholine has been proposed as a possible treatment for patients suffering from ulcerative colitis as it has been shown to have anti-inflammatory effects and is a component in the protective colonic mucus (Treede et al., 2009).

Protein Fraction. In addition to the MFGM phospholipids, some MFGM proteins have been shown to have anticarcinogenic activities. The BRCA1 and BRCA2 oncosuppressor proteins have been detected in human and bovine MFGM extracts (Vissak et al., 2002). Both the BRCA1 and BRCA2 proteins are involved in DNA repair processes and are known inhibitors of breast cancer. There is also evidence of several of the MFGM proteins contributing to the prevention of microbial infections. Wang et al. (2001) studied the inhibition of Helicobacter pylori infection by bovine MFGM components in a mouse model and noted that both the defatted and non-defatted MFGM preparations given orally caused equal healing effects on infection of the gastric mucosa.

Although many benefits of consuming MFGM proteins have been noted, some potential negative effects have also been suggested. Apparent links associating multiple sclerosis and autism with milk consumption and an immune response against the MFGM protein butyrophilin have been cited (Riccio, 2004); however, milk consumption is only one of the many environmental factors related to disease prevalence. Butyrophilin shows molecular mimicry with
myelin oligodendrocyte glycoprotein, a component of the myelin membrane that is a candidate autoantigen in human multiple sclerosis and may be capable of inducing experimental autoimmune encephalomyelitis (EAE), a disease displaying similar characteristics to multiple sclerosis, in experimental animals (Stefferl et al., 2000).

MILK FAT GLOBULE SIZE

Milk fat globules are secreted in a wide variety of sizes ranging in diameter from about 0.1 to 15 µm. Walstra (1969) reported MFG with a diameter less than 1 µm represented 80% of MFG, but only 5% of milk fat volume, while MFG ranging from 1 to 8 µm comprised approximately 94% of fat. The remaining 1 to 2% of the fat volume comes from few, large MFG greater than 8 µm in diameter. The average bovine MFG is approximately 2.5 to 4.6 µm in diameter although globule size shows considerable variation (Huppertz and Kelly, 2006).

Determining and Expressing Milk Fat Globule Size

Different analytical methods used to describe MFG size may yield different results, and results obtained from some older methods may be unreliable (Huppertz and Kelly, 2006). The size of MFG can be determined accurately by many more recent methods, including dynamic or static light scattering, Coulter counting, electroacoustics, ultrasonic spectroscopy and light or electron microscopy (Huppertz et al., 2009). During the measurement of MFG it is important to avoid the interference of other milk components. To avoid the interference of casein micelles, they can be dissociated with calcium-chelating agents (e.g., trisodium citrate or ethylenediamine tetra-acetic acid; EDTA) (Huppertz and Kelly, 2006). It may also be advisable to ensure that MFG are not clustering in the sample.

Mean MFG size can be represented using several parameters. Milk fat globule size is usually described by the diameter, irrespective of true MFG shape that may not be truly spherical. The ascertained diameter will be that of a sphere having equivalent properties (e.g. surface area or volume) to that of the particle. The most common parameters are derived from the ‘moments’ of the particle size distribution. The $n^{th}$ moment of a particle size distribution ($S_n$) is given by:

$$S_n = \sum N_i d_i^n,$$
where $N_i$ and $d_i$ are the number and diameter, respectively, of the particles in size class $i$ (Walstra, 2003).

Different moments of the particle size distribution, most commonly $S_1$, $S_2$, $S_3$, and $S_4$, can be used to derive parameters expressing average particle size. The number mean diameter ($D[1,0] = S_1/S_0$) and volume mean diameter ($D[3,1] = (S_3/S_1)^{1/3}$) can be reported; however, a large number of very small particles can dominate the result when they comprise a very small proportion of the total volume. The volume moment-weighted mean or De Brouckere mean diameter ($D[4,3] = S_3/S_3$) better reflects the size of the particles that constitute the majority of the sample volume. This parameter is sensitive to the presence of large particles in the size distribution. The volume surface-weighted mean or Sauter mean diameter ($D[3,2] = S_3/S_2$), is the weighted average surface diameter, assuming spherical particles of the same surface area as the actual particles. This measure is more relevant where specific surface area is meaningful such as with the MFGM. It is more sensitive to the presence of fine particles in the size distribution.

**Factors Influencing Milk Fat Globule Size**

The size distribution and average diameter of MFG depend on a variety of factors including breed, the individual animal, parity, stage of lactation, diet, and season (Mulder and Walstra, 1974; Walstra, 1995; Wiking et al., 2003; Carroll et al., 2006; Martini et al. 2013; Logan et al., 2014). Jersey and Guernsey breeds tend to give milk with larger MFG than Holstein-Friesian-type cows (Mulder and Walstra, 1974). Carroll et al. (2006) also found MFG from Jersey cows were larger in diameter on average than Holstein cows as well as Brown Swiss cows. Wiking et al. (2004) found a correlation between average MFG diameter and diurnal fat production for Danish Holstein cows. They theorized that MFGM material synthesis might not be able to increase as fat yield increases and as a result fat droplets grow larger before they are enveloped with plasma membrane in the secretory apical membrane. Larger MFG could be secreted to reduce the amount of membrane lost per unit volume of fat when membrane material is limited. According to Mulder and Walstra (1974), the diameter of MFG is at a maximum early in lactation and decreases throughout lactation. This decrease was also observed by Wiking et al. (2003) and is likely related to the correlation found between diameter and fat yield. The ability to modify MFG size by altering the diet of cows has been examined. The mean MFG size has been shown to decrease when a greater amount of fresh grass was included in the diet (Couvreur et al.,
Similar conclusions were made by Briard et al. (2003), who found smaller MFG in spring milk when cows had increased pasture feeding compared to winter milk. When cows were fed concentrate with a high quantity of saturated lipids, Wiking et al. (2003) found the milk had a greater fat content and significantly larger MFG than milk from cows fed other diets. This was also shown by Carroll et al. (2006) who observed, based on 12 cows from 3 dairy breeds, that average MFG diameter tended to increase with increased fat production, as a result of feeding diets with increased fat. Conversely, Walstra (1969) did not find average MFG to be correlated with fat content, nor a difference between cows or within breed (Friesian and Jersey).

The potential genetic regulation of MFG size was suggested by Campbell (1932). Logan et al. (2014) noted variation in MFG between individual cows within a herd of Holstein-Friesian cows. There is a variation in MFG diameter of up to 1 µm between individual animals (Mulder and Walstra, 1974) suggesting the possibility of selecting cows based on MFG size. Few studies have looked to investigate the genetic influence on MFG size. A 1965 study by Hassanein reported a heritability of 0.73 for butterfat globule size using a half-sib method in German black and white cows. The heritability of MFG size distribution was examined in Italian Friesian cows by Cabassi et al. (2013) who found estimates of 0.26 and 0.22 for surface mean diameter (D[3,2]) and volume mean diameter (D[4,3]), respectively. They also reported D[3,2] and D[4,3] to have a very high genetic correlation (0.98) with each other. Both size distribution measures had positive genetic correlations with milk yield, fat yield, and somatic cell score and a negative genetic correlation with protein yield. This study however only included 50 cows from 6 sires.

**Compositional Differences**

The main compositional differences in lipids between small and large MFG relate to the ratio of the MFG core to the membrane. The size of the MFG is a critical factor in the amount membrane material present. Small MFG have more membrane material per unit of fat than large MFG. The relative concentration of phospholipid, as well as other membrane materials, to triacylglycerol will therefore be higher with small MFG. Differences in the triacylglycerol core to MFGM ratio also indicate differences in FA composition. MFGM contain relatively larger quantities of unsaturated FA than the triacylglycerol core due to the FA composition of MFGM phospholipids, especially sphingomyelins (Jensen and Newburg, 1995; Fauquant et al., 2005). In MFG separated by microfiltration, Briard et al. (2003) discovered higher amounts of C18:1 and
C18:2 in small MFG than in large MFG and concluded the increased amounts were more than could be attributed to increased MFGM alone.

Wiking et al. (2004) reported the average diameter of MFG was positively correlated with C16:0, C16:1, C18:0 and C18:1 but no correlation with C4:0 to C14:0, C18:2, or C18:3. Since many of these FA originate from the diet, this study concluded that diet is a factor in MFG size and it affects the compositional differences. Lopez et al. (2011) did not find differences in the amount of total saturated and unsaturated fatty acids or in the amounts of short chain fatty acids (C6:0 to C10:0) between different sized MFG. They did, however, find that in small MFG there were significantly more C12:0, C14:0 C16:0, C18:1 trans, and C18:2c9t11, and less C18:0 and C18:1c9 than in large MFG. Conversely, Timmen and Patton (1988) found less small-chain fatty acids (C4:0 to C10:0) in small MFG compared to large MFG, but did also find less C18:0 and greater C18:1. The conflicting results can likely be attributed to large differences in methods used between these studies.

The composition of MFG core and membrane were examined separately by Fauquant et al. (2005) after using cross-flow microfiltration to divide MFG by size. They found no significant differences in FA composition between small and large MFGM, which gives further evidence that compositional differences between different sized MFG cannot be explained by the MFGM alone. In the triacylglycerol core, they detected significantly more C12:0, C14:0, C14:1, C16:0, C16:1, C21:0, and C20:3 n-3 and less C18:0 and C20:3 n-3 in small MFG compared to large MFG.

Compositional differences in the MFGM have also been examined for changes in the types of polar lipids with MFG size. Lopez et al. (2011) noted higher amounts of polar lipids per gram of fat in small MFG fractions than large MFG fraction, which was expected due to increased MFGM material. They also found small MFG fractions contained significantly lower relative proportions of phosphatidylcholine and sphingomyelin compared to large MFG fractions and whole milks. In contrast, Mesilati-Stahy et al. (2011), who also examined phospholipid differences in the MFGM, found the smallest MFG fraction had more sphingomyelin by percent weight than the largest MFG fraction and found no relationship between phosphatidylcholine concentration and MFG size. This study, however, separated MFG into 6 size groups using a gravity-based method, which is not as accurate, and not all of the size groups had significantly
different mean diameters. No significant differences in phospholipid composition between different sized MFG within a herd were detected by Logan et al. (2014).

Mulder and Walstra (1974) cautioned that results of compositional differences found in bulk milk might be inadvertently detecting cow-to-cow differences. However, it has been found that the compositional differences of MFG of equal size from a single milking of one cow can be similar in magnitude to those found between cows (Walstra and Borggreve, 1966). It is also important to note that the nature and magnitude of compositional changes observed between small and large MFG can be influenced by the diet (Wiking et al., 2004; Lopez, 2008), breed (Gallier et al., 2011), and season (Briard et al., 2003).

Effect on Milk Processing and Products

Milk fat globule size has critical implications for the technological and sensory properties of many milk products. These differences may be related to variation in their composition or changes in the way the MFG exist in and interact with their environment. As mentioned before, milk fat has a broad melting range because of their very diverse composition and configuration, and the melting point of triacylglycerols depends on their fatty acid composition. Long-chain fatty acids have a higher melting point than short-chain fatty acids as well as saturated fats over unsaturated fats. The differences found between MFG of different sizes, in terms of fatty acid composition, could therefore cause changes in the melting characteristics of the milk fat, which has implications for its technological functionality.

Creaming rate, the proportion of fat arriving in the cream layer per unit of time, is affected by mean MFG size, as large globules will rise faster (Huppertz and Kelly, 2006). Cream with small MFG churns more slowly and more globules may escape during the churning process producing more fat in buttermilk than cream with large MFG (Mulder and Walstra, 1974). Moreover, smaller MFG may adversely affect crystallization partly on account of more disordered crystallization and smaller fat crystals (Huppertz and Kelly, 2006), harming butter yield, viscosity, elasticity, and product consistency.

Milk fat globules can be separated by size without damaging the MFGM using membrane microfiltration to study differences in dairy products created from small or large MFG (Goudédranche et al., 2000). During the manufacturing process of Camembert cheeses produced using small MFG, Michalski et al. (2003) found less whey was collected compared to cheeses
produced using large MFG. This ultimately led to higher moisture content in the ripened cheeses and more fat retained. Additionally, small MFG Camembert cheeses had higher elastic and melting texture, and higher flowing aspect while large MFG cheeses had a more firm and chalky texture and yellower colour. Parallel findings were also reported for Emmental cheese produced with different sized native MFG, where small MFG cheese retained more moisture and were less firm and flexible (Michalski et al., 2004). These differences are explained by the fact that at an equal fat content, milk made up of small MFG will have a greater number of MFG with a larger surface area of MFGM, with a higher water-binding ability (Michalski et al., 2003).

Effects of Homogenization

Homogenization is a process designed to disrupt MFG in milk to make them much smaller, thereby changing the fat dispersion and preventing creaming of the fat. The mean diameter of the MFG decreases to less than 1 µm after homogenization (Keenan and Patton, 1995). The resultant MFG size distribution depends on the mechanism of homogenization, the amount of pressure used, and the properties of the milk (Mulder and Walstra, 1974). Smaller, and thus more MFG, are created with a greater surface area, but with a changed membrane. The original amount of MFGM material is not enough to cover and stabilize the new large surface area. The newly exposed triacylglycerol adsorb milk proteins, particularly casein micelles (Keenan and Patton, 1995). The disrupted MFG are covered mainly in caseins and some of the original MFGM material, forming lipid-protein complexes. Some very small native MFG may be unaffected by homogenization and retain their native MFGM (Michalski et al., 2002). The new membrane structures that are formed differ in composition and give the MFG different physical and chemical properties (Michalski et al., 2004). For this reason, products with small MFG created by the homogenization process are not comparable with products with small native MFG.

INFRARED SPECTROSCOPY

Infrared (IR) refers to the part of the electromagnetic spectrum between the visible and microwave regions and is subdivided into the near-, mid-, and far-IR regions. The mid-infrared (MIR) region (2,500-25,000 nm or 4,000-400 cm⁻¹) is the main region of vibrational spectroscopy, which allows for the identification of organic molecules and characterization of their structure and composition (Dufour, 2009).
In IR spectroscopy, IR radiation is applied to the sample and absorbed by molecules creating vibrational movements. In the MIR range absorption is generally from the vibrational ground state to the first excited vibrational state. Absorption generally occurs when the frequency of the IR equals the frequency of the molecular vibration and when the molecular dipole moment changes during the vibration. The vibrational frequency and probability of absorption depend on the strength and polarity of the bond. The type of bond, the exact position of electron withdrawing or donating, effects of the intra- and intermolecular environment, and coupling with other vibrations determine the approximate position of the absorption band (Barth, 2007). The amount of transmitted light is examined to determine how much energy was absorbed at each frequency. The intensity of the band at a given wavelength can be expressed as percent transmittance, where zero transmittance corresponds to 100% absorption of light and absorption of radiant energy is represented by a trough in the curve, or as absorbance, the logarithm, to the base 10, of the reciprocal of the transmittance, where a peak in the curve signifies the absorption of energy. Transmittance and absorption are plotted against wavelength (nm) or wavenumber (cm⁻¹).

Identification and the attribution of these bands to specific chemical groups give specific information on the investigated product. In addition, the absence of important bands can also be useful in the interpretation of the MIR spectra. The MIR region can be generalized into four regions: the X-H stretching region (4,000-2,500 cm⁻¹), the triple-bond region (2,500-2,000 cm⁻¹), the double-bond region (2,000-1,500 cm⁻¹), and the fingerprint region (1,500-600 cm⁻¹) (Stuart, 2004). In the 4,000-2,500 cm⁻¹ range, O-H stretching produces a broad band between 3,700 and 3,600 cm⁻¹, N-H stretching between 3,400 and 3,300 cm⁻¹, and C-H stretching between 3,000 and 2,850 cm⁻¹ for aliphatic compounds and between 3,100 and 3,000 cm⁻¹ if the bond is adjacent to a double bond or aromatic ring (Stuart, 2004). In the triple bond region, the most common bonds are C≡C that absorbs between 2,300 and 2,050 cm⁻¹ and is normally very weak, and C≡N that absorbs between 2,300 and 2,200 cm⁻¹ at a medium intensity (Stuart, 2004). It is also possible for some X-H stretching absorptions in this region when X is a very large atom. One of the most intense bands in the MIR spectrum is C=O which is usually seen in 1,830-1,650 cm⁻¹ depending on the type of bond, but may absorb above 2,000 cm⁻¹ for metal carbonyls (Stuart, 2004). C=C and C≡N bonds absorb at approximately 1,650 cm⁻¹ but C≡C is much weaker and often absent for symmetry or dipole moment reasons (Stuart, 2004). The assignment of many more specific
spectral band frequencies by functional group and bond type is given by Coates (2000).

In MIR milk recording, all of the molecules in the milk sample interact with the MIR laser and thus the MIR spectrum represents the global composition of milk. Each milk component has its own absorption profile across the MIR range, and the absorption profile of the complete milk sample is an aggregation of all of these component specific profiles. MIR analysis is used to measure fat, protein, and lactose in milk samples with high accuracy. Historically, prediction of these traits was done with fixed filters using four primary bands: fat A (carbonyl stretch; 1,745 cm\(^{-1}\)), fat B (alkyl stretch; 2,874 cm\(^{-1}\)), amide II (amide stretch; 1,548 cm\(^{-1}\)), and lactose (hydroxyl stretch; 1,040 cm\(^{-1}\)). Recent Fourier Transform Infrared (FT-IR) technology in spectroscopy produces the full spectral information within the MIR range simultaneously, increasing the speed and sensitivity of spectral machines. This provides an increased quantity of information allowing the quantity of other milk components to be predicted.

**Challenges of MIR and Milk**

Milk is a very complex substance with many different components and thus has a very complex spectrum. The MIR spectral range itself is very intricate and it can be difficult to assign bands to individual chemical groups because many bands will have contributions from the vibrations of several chemical groups. The process also needs to be very controlled as any variation in sample preparation can affect the predictions.

Mid-infrared radiation has limited penetration depth and this causes the spectra to be very sensitive to the presence of MFG or fat biofilms. Consequently, MIR spectroscopy is not suitable for on-line, real time raw milk analysis (Linker and Etzion, 2008). This issue is solved by the use of a homogenizer but this limits the use of MIR analysis of milk to manipulated milk offline.

Water presents one of the primary challenges in using MIR spectroscopy. Because milk is chiefly comprised of water, it makes a large contribution to the infrared absorption. Water absorption is very strong in the infrared region with prominent bands centred at approximately 1,640 cm\(^{-1}\) (O-H scissoring) and 3,300 cm\(^{-1}\) (O-H asymmetric and symmetric stretching). The domination of water at these bands can obscure lesser protein bands like amide I (1,655 cm\(^{-1}\)) and, to a lesser extent, amide II (1,560 cm\(^{-1}\)). The high absorption of water in these bands also creates a low signal to noise ratio thereby making information largely unattainable.
The predicted values for milk traits using MIR will only ever be as good as the methods used to create the calibration set for developing prediction equations. Highly accurate values for parameters are required in order to create accurate equations for their prediction. Any error in the analysis of the calibration set samples will create error in the final predictions using MIR. Therefore, it is vital to have a gold standard test as close as possible to the hypothetical ideal gold standard test for the analysis of component traits to be predicted with MIR.

**Genetic Variability of the Spectra**

The MIR spectrum of milk represents its composition and thus genetic variation of milk components should be reflected as genetic variability of the MIR milk spectrum. Soyeurt et al. (2010) examined the genetic variability of the MIR milk spectral data by first performing a principal components analysis and creating 46 new traits from 46 principal components. Heritabilities of the new traits ranged from 0.00 to 0.35 with eight having heritabilities greater than 0.10. Variances of the original spectral traits were obtained by back transformation and their heritabilities ranged from 0.003 to 0.42. Regions of the spectra located between 926 and 1,612 cm\(^{-1}\), 1,682 and 3,062 cm\(^{-1}\), and 3,672 and 5,010 cm\(^{-1}\) were identified as having moderate to high genetic variability. Bittante and Cecchinato (2013) also discovered a direct relationship between the cow’s genetics with 1,056 transmittance data points of the MIR spectra of her milk by estimating heritabilities for each spectral point individually. They found regions of the MIR spectra associated with important chemical bonds have higher and less variable heritability estimates than regions associated with water absorption.

**PREDICTING NOVEL TRAITS USING MIR**

The number of traits being predicted from the MIR spectra of milk has been rapidly growing recently, as reviewed by De Marchi et al. (2014). This includes the prediction of phenotypes for more detailed milk components, such as fat and protein composition and mineral contents, milk technological properties, and cow physiological status.

Common measures used to report the accuracy and utility of the calibration equations are the coefficient of determination (\(R^2\)), the ratio of performance deviation (RPD), and the range error ratio (RER). The \(R^2\) of a model is the amount of the total response variation explained by the explanatory variable, and is calculated as one minus the quotient of the residual sum of
squares divided by the total sum of squares of the model. Soyeurt et al. (2011) proposed that prediction equations with an $R^2$ in validation greater than 0.95 could be used for payment purposes, while equations with an $R^2$ in cross validation greater than 0.75 could be used for animal breeding purposes. However, there may still be utility in traits with predicted values below this suggested threshold (Cecchinato et al., 2009). The RPD is calculated as the ratio of standard deviation of the calibration set to the standard error of cross-validation (SECV). A higher value is desired, with a RPD greater than 2 enabling good predictions (De Marchi et al., 2011). The RER is calculated by dividing the range of the reference data by the SECV. Calibration equations with a RER value between 7 and 20 are considered poor to fair and may be adequate for screening purposes, and values between 21 and 40 are good to very good and more suitable for quality and process control applications (Williams, 2001).

The goal in creating a prediction equation is to model the correlation between the spectrum and the trait. Partial least squares (PLS) regression is an attractive and widely used method for formulating MIR predictive models due to the large number of correlated data points obtained as part of the milk MIR spectra. Partial least squares regression works by defining a set of latent variables, accounting for as much of the covariance between the predictors and responses. Ferrand-Calmels et al. (2014) found PLS methods outperformed least absolute shrinkage and selection operator (LASSO), and elastic net regression methods. Ferragina et al. (2015) suggested the use of Bayesian methods (Bayesian ridge regression, Bayes A, and Bayes B) for deriving calibration models over PLS. They found prediction accuracies were significantly higher using Bayesian methods than PLS methods for their selected individual milk FA and milk technological properties, with Bayes B generally performing the best. The $R^2$ values for the predicted FA reported by Ferragina et al. (2015) for their PLS models were lower than those reported by other authors (e.g. Soyeurt et al., 2011), so it will have to be seen if Bayesian methods can provide improvements when used with these calibration sets which already perform well.

Statistical procedures for pre-processing spectral data, including scatter correction and derivation methods, can be used to improve the linear relationship between the spectra and the reference values (De Marchi et al., 2014). Nonlinearities from particulates in the sample can be taken out using multiplicative scatter correction. Derivatives can be used to help with overlapping absorption bands and large baseline variations (Hruschka, 2001). Soyeurt et al.
(2011) compared derivative pre-treatments on the spectra for the accuracy of individual FA predictions and found first derivative pre-treatments yielded the best results for the majority of FA. De Marchi et al. (2011) also reported most FA predicted better when the first derivative or first derivative plus multiplicative scatter correction was performed on the spectra prior to PLS, compared to no spectral pre-treatment. The same was not found by De Marchi et al. (2013) for milk coagulation properties and McParland et al. (2011) for energy status, as no significant improvement was found by using spectral pre-treatments. Most studies do not report the results for all the pre-treatments attempted and thus their effect is hard to evaluate.

The elimination of uninformative variables, such as noisy or random variables, can improve the predictive ability of multivariate calibration models. Niero et al. (2016) used an uninformative variable elimination procedure followed by PLS to generate MIR calibration models for milk protein composition, reducing the number of predictors from 865 down to 110 to 390, depending on the trait, and noted improvements in accuracy for all traits.

The properties of the calibration set of gold standard measured data are important for the utility of the final prediction model. The variation in the calibration set needs to be as great as possible for the PLS procedure to work and the calibration set should cover the expected variation in the population to be predicted (McParland et al., 2011; De Marchi et al., 2014). Thus, the sample size needs to be sizeable enough to capture as much variation as possible in both the predictors and response variables. Predictions of milk minerals were improved with calibration sets with higher variability and where more extreme values were included, as presented by Soyeurt et al. (2009). Rutten et al. (2010) examined the impact of differently sized calibration sets for the MIR prediction of some FA and found a strong relationship between the number of samples and the validation $R^2$ as they increased the sample numbers from 100, 250, 500, and 1,000. They suggested that there is likely a threshold for the number of calibration samples from which no further improvement in $R^2$ can be achieved.

**Predicting Milk Components**

**Fatty acid composition.** The prediction of milk FA contents using MIR spectroscopy has been pursued by several authors (Soyeurt et al., 2006a; Rutten et al., 2009; De Marchi et al., 2011; Ferrand et al., 2011; Soyeurt et al., 2011; Maurice-Van Eijndhoven et al., 2013; Ferrand-Calmels et al., 2014). The large range in observed results are due to the studies employing different
populations, numbers of samples, spectra pre-treatments, reference methods, and different units for expressing FA (De Marchi et al., 2014). Soyeurt et al. (2006a) and Rutten et al. (2009) observed higher accuracies when fatty acids were expressed on a per milk basis versus per fat basis.

The accuracy of the prediction of individual FA greatly varies between the different FA. The lowest $R^2$ in cross-validation identified by Soyeurt et al. (2006a) for an individual FA was 0.01 for C10:1c9 and the greatest was 0.88 for C18:1. In total, Soyeurt et al. (2011) were able to predict 10 individual FA using models with $R^2$ in cross-validation greater than 0.80. Rutten et al. (2009) was also able to predict individual even-chain FA from C4:0 to C18:0 and C18:1c9 all with $R^2$ above 0.80. Overall, lower accuracies were observed by Soyeurt et al. (2006a) and De Marchi et al. (2011) for predicting FA, possibly due to the use of smaller and less diverse calibration sets. The FA appearing in the greater concentrations in milk tend to predict with the greatest accuracies using MIR data. The positive relationship between FA concentration and predictive model performance has been identified and discussed by Soyeurt et al. (2006a) and De Marchi et al. (2011). Rutten et al. (2009) modeled the relationship between fatty acid concentration (g/dL) and prediction $R^2$ and reported an $R^2$ value for this relationship of 0.64.

Groups of FA generally predict well for all authors. Soyeurt et al. (2011) had $R^2$ in cross-validation greater or equal to 0.95 for saturated, monounsaturated, unsaturated, and short-, medium-, and long-chain FA. Only polyunsaturated FA did not predict as well ($R^2 = 0.70$). Ferrand-Calmels et al. (2014) also found the polyunsaturated FA group to predict much lower than saturated and monounsaturated FA groups. This is likely due to their lower concentration. Rutten et al. (2009) reported high $R^2$ values of 0.95, 0.97, and 0.94 for short-, medium-, and long-chain FA, respectively.

**Protein composition.** Protein composition typically does not predict as well as fat composition (De Marchi et al., 2014). Bonfatti et al. (2011) investigated the ability to predict detailed casein and whey protein components using MIR spectra and found generally low accuracies. Better results were found when components were expressed as g/L of milk compared to percent of protein or caseins, much like what was found for FA composition. The coefficients of determination of calibration ranged from 0.10 to 0.69 for the casein fractions (g/L) and from 0.39 to 0.67 for examined whey fractions (Bonfatti et al., 2011). Similar, unsatisfactory results were
also reported by McDermott et al. (2016). Although models poorly predict protein composition, Bonfatti et al. (2011) suggested the predicted values might still have a benefit for genetic selection. The β-Lactoglobulin genotype is significantly associated with the composition of bovine milk protein and Rutten et al. (2011) showed that using spectral data one could correctly identify the β-Lactoglobulin genotype of a cow on average 74% of the time.

Soyeurt et al. (2012) looked to predict the glycoprotein lactoferrin in milk from spectral data and achieved R² in cross-validation ranging between 0.69 and 0.73. This trait would be of particular interest as it is a potential indicator of mastitis, and when used alongside somatic cell score, may improve mastitis detection (Soyeurt et al., 2012).

**Other milk components.** The prediction of milk calcium, potassium, magnesium, sodium, and phosphorus mineral contents using MIR spectroscopy was attempted by Soyeurt et al. (2009), and the reported R² in cross-validation for the minerals were 0.87, 0.36, 0.65, 0.65, and 0.85, respectively. The predictions of calcium and phosphorus developed were related to the content of fat in milk, which may explain their high accuracy. The mineral contents were correlated with test-day fat content and the spectral regions related to fat absorption were highly utilized by the predictive models.

Hansen (1999) reported an R² of 0.81 for the prediction of acetone from milk spectral data. The prediction was also tested on samples fortified with acetone and low accuracies were determined, suggesting it was not just the relationship between acetone and the spectra being modeled. The predictions for acetone were not powerful enough to accurately identify cows suffering from ketosis (Hansen, 1999). Following these results, Heuer et al. (2001) limited the spectral variables to those surrounding wave numbers related to acetone and found improved results. They also employed samples spiked with acetone and found these samples cannot be used in the production of the prediction equations because of a possible relationship between acetone, other ketone bodies, and milk fat in natural milk samples. Using both MIR-predicted acetone and BHB may be more useful in detecting ketosis in dairy cows (de Roos et al., 2007).

**Predicting Technological Properties**

The predictions of many more complex traits from MIR spectra related to cheese production have been examined. Cecchinato et al. (2009) reported R² values of 0.61 to 0.69 for
rennet coagulation time and 0.46 to 0.52 for curd firmness at 30 minutes. The prediction of milk coagulation properties, rennet coagulation time, curd-firmness time, and curd firmness, have also been presented by De Marchi et al. (2013), Chessa et al. (2014), and Visentin et al. (2015). De Marchi et al. (2013) had the greatest $R^2$ values for rennet coagulation time, curd firming time, and curd firmness at 30 minutes at 0.82, 0.80, 0.87, respectively. This may be a result of many factors including the samples and methodologies used. Visentin et al. (2015) reported prediction results for the additional traits of heat coagulation time, casein micelle size, and pH with validation $R^2$ values of 0.46, 0.13, and 0.71, respectively. Calamari et al. (2016) were able to accurately predict titratable acidity in milk samples with a wide variety of milk characteristics with a model $R^2$ of 0.96. The prediction models for titratable acidity were related to the areas of the spectra for protein content and carboxylic acid or esters, which are known to be factors involved in milk titratable acidity (Calamari et al., 2016).

**Predicting Cow Systemic Traits**

McParland et al. (2011) attempted to use MIR data to predict direct energy balance, body energy content, body condition score, and effective energy intake of Holstein-Friesians in Scotland at different stages of the lactation and milking times. Effective energy intake was predicted the best with correlation coefficients for external validation ranging between 0.68 and 0.87. The greatest correlation coefficient achieved for direct energy balance was 0.75 when using evening milk samples across lactation. The inclusion of milk yield in the predictive models was able to increase the correlation coefficient for effective energy intake by 0.15 to 0.18 units. However, these equations can only be used to predict the energy status of animals in a population with variation well represented by the calibration set, and are thus not robust enough to be used in different production systems or populations with dissimilar energy status (McParland et al., 2011, 2012). McParland et al. (2014) predicted the added trait of residual feed intake and reported coefficients of correlation from 0.48 to 0.60 in external validation. Building on these previous results, Hempstalk et al. (2015) tried various machine learning algorithms to predict the likelihood of conception from the milk MIR spectra of dairy cows as it is a trait related to energy balance and body condition score change, which can be predicted from the MIR spectrum. They concluded that MIR did not offer any improvement in predicting conception outcomes above other routinely available factors.
The prediction of methane production using milk MIR spectra of lactating dairy cows has also been attempted, mostly indirectly using MIR-predicted FA contents. This application may be limited since many FA related to methane production do not predict well from MIR spectra (van Gastelen and Dijkstra, 2016). Dehareng et al. (2012) looked to directly predict methane production traits from MIR spectral data and found high R$^2$ in cross-validation ranging from 0.68 to 0.79. Vanlierde et al. (2015) stated that these results were explained by the lactation stage, and thus, they also developed a prediction equation including days in milk to obtain a more biologically meaningful prediction of methane. Their R$^2$ of calibration for the prediction equations independent and dependent of lactation stage were 0.77 and 0.75, respectively.

**SUMMARY**

The fat component of bovine milk shows variation in its composition and structure. Depending on the FA chain length, milk FA are either synthesized *de novo* in the mammary gland or provided directly to the mammary gland from the bloodstream as preformed FA. In early lactation, when the cow is in negative energy balance, there is increased mobilization of body reserves, and consequently increased proportions of long-chain FA in milk. Fatty acid contents could therefore be used as an indicator for energy balance, which can have repercussions for cow health. The incorporation and quantity of various FA into milk can also be impacted by the diet and the genetics of the cow. Fatty acid contents in milk are heritable and, therefore, potential traits for selection.

The mean size of MFG in milk can differ by breed, parity, lactation stage, and diet. The mechanisms for the determination of MFG size, however, are not completely understood. Variation in MFG size has been found between individual cows. Therefore, there may be a genetic component to MFG size that has not been fully investigated yet. Milk fat globule size is of interest due to the health benefits of the membrane materials and its effects on the production of milk products, like cheese.

The MIR spectrum of milk is a reflection of the global composition of a sample. Mid-infrared spectroscopy can be effectively used to predict a multitude of new traits related to milk composition and properties, and cow traits, making the traits available for a large number of animals. The ability to predict the quantity of many fatty acids in milk has been shown to be possible with variable accuracy in several European countries. This technology could also be
applied to the Canadian population to obtain FA phenotypes for further research. The prediction of MFG size using MIR spectroscopy has not been undertaken previously. Milk fat compositional differences have been noted in milk with differently sized MFG. The major difference is the quantity of triacylglycerides of the core versus the phospholipids of the membrane. Differences in fatty acids have also been found, but these may not be constant across populations or management practices. Mid-infrared spectral data may hold information on the compositional differences related to MFG size, and, therefore, prediction of mean MFG size may be achievable.
Table 2.1. Main fatty acids by percent weight of total in triacylglycerols in bovine milk (Christie, 1995).

<table>
<thead>
<tr>
<th>Fatty Acid</th>
<th>Common name</th>
<th>% Weight of Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>C4:0</td>
<td>Butyric acid</td>
<td>3.1</td>
</tr>
<tr>
<td>C6:0</td>
<td>Capronic acid</td>
<td>1.6</td>
</tr>
<tr>
<td>C8:0</td>
<td>Caprylic acid</td>
<td>1.3</td>
</tr>
<tr>
<td>C10:0</td>
<td>Caprinic acid</td>
<td>3.0</td>
</tr>
<tr>
<td>C12:0</td>
<td>Lauric acid</td>
<td>3.1</td>
</tr>
<tr>
<td>C14:0</td>
<td>Myristic acid</td>
<td>9.5</td>
</tr>
<tr>
<td>C16:0</td>
<td>Palmitic acid</td>
<td>26.3</td>
</tr>
<tr>
<td>C16:1</td>
<td>Palmitoleic acid</td>
<td>2.3</td>
</tr>
<tr>
<td>C18:0</td>
<td>Stearic acid</td>
<td>14.6</td>
</tr>
<tr>
<td>C18:1</td>
<td>Oleic acid</td>
<td>29.8</td>
</tr>
<tr>
<td>C18:2</td>
<td>Linoleic acid</td>
<td>2.4</td>
</tr>
<tr>
<td>C18:3</td>
<td>Linolenic acid</td>
<td>0.8</td>
</tr>
<tr>
<td>C20-C22</td>
<td>Trace</td>
<td></td>
</tr>
</tbody>
</table>

Table 2.2. Gross composition of cow milk fat globule membrane (Keenan and Patton, 1995).

<table>
<thead>
<tr>
<th>Constituent class</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proteins</td>
<td>25-60 g/100g</td>
</tr>
<tr>
<td>Total lipids</td>
<td>0.5-1.1 mg/mg protein</td>
</tr>
<tr>
<td>Phospholipids</td>
<td>0.13-0.34 mg/mg protein</td>
</tr>
<tr>
<td>Neutral lipids</td>
<td>0.25-0.88 mg/mg protein</td>
</tr>
<tr>
<td>Glycosphingolipids</td>
<td>13 µg/mg protein</td>
</tr>
<tr>
<td>Hexoses</td>
<td>108 µg/mg protein</td>
</tr>
<tr>
<td>Hexosamines</td>
<td>66 µg/mg protein</td>
</tr>
<tr>
<td>Sialic acids</td>
<td>20 µg/mg protein</td>
</tr>
<tr>
<td>Glycosaminoglycans</td>
<td>0.1 µg/mg protein</td>
</tr>
<tr>
<td>RNA</td>
<td>20 µg/mg protein</td>
</tr>
</tbody>
</table>

Table 2.3. Lipids of the milk fat globule membrane (Keenan and Patton, 1995).

<table>
<thead>
<tr>
<th>Lipid Class</th>
<th>% of total lipid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Triacylglycerols</td>
<td>62</td>
</tr>
<tr>
<td>Diacylglycerols</td>
<td>9</td>
</tr>
<tr>
<td>Monocacylglycerols</td>
<td>0-0.5</td>
</tr>
<tr>
<td>Unesterified fatty acids</td>
<td>0.6-6.0</td>
</tr>
<tr>
<td>Phospholipids</td>
<td>26-31</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lipid Class</th>
<th>% of total phospholipid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sphingomyelin</td>
<td>22</td>
</tr>
<tr>
<td>Phosphatidylcholine</td>
<td>36</td>
</tr>
<tr>
<td>Phosphatidylethanolamine</td>
<td>27</td>
</tr>
<tr>
<td>Phosphatidylinositol</td>
<td>11</td>
</tr>
<tr>
<td>Phosphatidylserine</td>
<td>2</td>
</tr>
</tbody>
</table>
Chapter 3

PREDICTION OF MILK FATTY ACID CONTENT WITH MID-INFRARED SPECTROSCOPY USING DIFFERENTLY DISTRIBUTED CALIBRATION SETS

ABSTRACT
The fatty acid profile of milk is a prevailing issue due to the potential negative or positive effects of different fatty acids to human health and nutrition. Mid-infrared spectroscopy can be used to obtain predictions of otherwise costly fatty acid phenotypes in a widespread and rapid manner. The objective of this study was to evaluate the prediction of fatty acid content for the Canadian dairy cattle population from mid-infrared spectral data and to compare the results produced by altering the calibration set used. The calibration sets used to calibrate the predictions were reference fatty acids expressed as (1) g/100g of fatty acid, (2) g/dL of milk, (3) the natural logarithmic transform of g/dL, and (4) subsets of samples randomly selected by removing excess records around the mean to present a more uniform distribution, repeated five times. Gas chromatography measured fatty acid concentration and spectral data for 1,958 milk samples from 372 cows from 4 breeds and 44 herds were used in the calibration. The coefficient of determination of cross-validation ($R^2_{cv}$) increased when fatty acids were expressed as g/dL compared to g/100g for all but C18:2n6t, which had very little relationship with the spectra. The logarithmic transformation used to create a more Gaussian distribution in the calibration had little effect on the predictions. The individual fatty acids C12:0, C14:0, C16:0, C18:0, C18:1n9c, and saturated, monounsaturated, unsaturated, short-chain, medium-chain, and long-chain fatty acid groups had $R^2_{cv}$ greater than 0.70. When calibration was performed with subsets of the original samples, slight increases in $R^2_{cv}$ values were observed for the majority of fatty acids. The difference in $R^2_{cv}$ between the top- and bottom-performing prediction equation across the different subsets for a single predicted fatty acid was on average 0.03 depending on which samples were randomly selected to be used in the calibration set. Predictions for fatty acids with high accuracies can be used to monitor fatty acid contents for cows in milk recording programs and possibly for genetic evaluation.

Key words: mid-infrared spectroscopy, fatty acid, calibration
INTRODUCTION

Milk and milk products are major contributors of nutrients to the diet of many humans and consumer awareness of the health effects of milk requires an increased look at its fatty acid profile. Over 400 different fatty acids have been identified in milk fat, although most appear only in trace amounts (Christie, 1995). In bovine milk fat there are only about 12 fatty acids present at above 1% concentration (Jensen, 2002). Milk fat contains approximately 70% saturated fatty acids. These fatty acids have been associated with increased levels of cholesterol and increased risk of cardiovascular disease, although some of the evidence supporting this claim may have alternate explanations (Pardi, 2004). In contrast, there is also evidence that some saturated fatty acids in milk may have neutral or positive effects on health and that milk fat is rich in many fatty acids beneficial to human health (Haug et al., 2007). Therefore, it may be of importance to alter fatty acid concentrations and optimize them for human health.

Beyond the nutritional aspect of fatty acids, the relative concentrations of different fatty acids have implications on the technological properties of milk and milk products (Huppertz and Kelly, 2009). As well, changes in milk fatty acids may be an indicator of cow health and energy status (Stoop et al., 2009). Therefore, strategies for managing or altering the fatty acid content of milk is of interest to the dairy industry, and a practical method for phenotyping this trait is required.

Milk composition testing is routinely performed by certified milk recording laboratories using mid-infrared (MIR) spectroscopy for payment, quality control, herd management, and genetic selection purposes. The obtained spectra of milk samples are used to simultaneously provide information on a variety of compositional parameters in a rapid and inexpensive manner. In 2015, approximately 700,000 cows in Canada (72% of all Canadian dairy cows) were enrolled on official or management milk recording programs (Canadian Dairy Information Centre). A high volume of data is produced and stored in the form of MIR spectra, which can potentially be exploited to predict additional milk traits. De Marchi et al. (2014) reviewed the expanded use of MIR for milk phenotyping. Mid-infrared spectra-based prediction of fatty acid content in milk has shown some, though varied, success using calibration sets from other dairy cattle populations (Rutten et al., 2009; De Marchi et al., 2011; Soyeurt et al., 2011; Lopez-Villalobos et al., 2014). Therefore, MIR technology may provide a method of obtaining fatty acid composition
phenotypes on large numbers of samples in Canada at a low cost from existing milk recording data.

The samples included in the calibration sample set have major implications on the effectiveness of the model in predicting new records. Selecting samples based on spectral characteristics has been a popular method in infrared calibration. This procedure involves recording the spectra, selecting samples likely to provide the best calibration based on the spectral variance, and finally performing reference analysis on these samples only. Mid-infrared predictions of fatty acids performed by Soyeurt et al. (2011) identified samples for their calibration set by first examining the spectra, which allowed for analyzing fewer milk samples yet still achieving a high amount of variation. In the present study, milk samples and MIR spectra were obtained during routine milk recording and the high volume of samples, the fast throughput of this process, and the vast geography involved, required samples marked for minor milk constituent analysis to be selected prior to collection. As well, further work with the recorded milk data is suited to multiple samples per cow. As a result of the sampling technique, the calibration sample sets for the fatty acids may have an overabundance of samples with the same composition. Restricting the number of similar samples included in the calibration set could better the prediction of some fatty acids.

The objective of this study was to investigate the capability of predicting fatty acid content from MIR spectral data collected during routine Canadian DHI milk recording by altering the calibration data set by adjusting the scale and distribution of samples in the calibration data set.

**MATERIALS AND METHODS**

*Milk Sampling*

Milk samples were collected during routine Canadian DHI milk recording by CanWest DHI (Guelph, ON) from February 2014 to October 2015 and Valacta (Sainte-Anne-de-Bellevue, QC) from February 2014 to May 2015. There were 44 participating herds located in the provinces of Alberta, Ontario, and Quebec with Ayrshire, Brown Swiss, Holstein, or Jersey breeds. From each herd, approximately 10 cows were identified (5 at the beginning and 5 at the middle part of the lactation on the first test) and multiple milk samples through one or two lactations were collected over the study period (20 months for ON and AB herds, and 15 months
for the QC herds). Individual cow milk samples (50 mL) were collected and sent to a DHI laboratory as per normal milk recording procedures. At the laboratory, the required quantity of milk needed for DHI milk testing was removed and the remainder was sent to the University of Guelph (Guelph, ON) for additional, fine milk component analysis.

**Milk Analysis**

Milk MIR spectra were obtained from one of two MilkoScan FT6000 spectrometers (FOSS, Hillerød, Denmark) at either CanWest DHI or Valacta laboratories following routine milk recording methodology. The MIR data for each sample contained 1,060 data points in the infrared range of 5,000 to 900 cm\(^{-1}\). Standardization of the historical spectra between the two machines and across time was performed per Bonfatti et al. (2016). For the purpose of creating prediction equations, regions 3,105 to 3,444 cm\(^{-1}\) and 1,628 to 1,658 cm\(^{-1}\) of the MIR spectra were removed due to low signal to noise ratio caused by the high absorption of water.

Milk fat extraction was performed at the University of Guelph using methods adapted from Chouinard et al. (1997) and methylation followed Christie's (1982) methods. Fatty acid composition was determined using an Agilent Technology model 7890B gas chromatograph (New Castle, DE, USA) equipped with an automatic on-column injector (Agilent G4513A), and a flame-ionization detector (FID) used on a CP-Sil88 fused silica capillary column (CP 7489 100m x 0.25mm x 0.2µm film thickness, Agilent J&W, USA). Column conditions were set up as follows. Methylated samples were dissolved in hexanes. Hydrogen was used as the carrier gas at a flow rate of 1mL/min. Then, 1µl of sample was injected directly cold on-column at an oven temperature of 35°C. After initiation, the column temperature was held at 35°C for 5 min, increased by 14°C/min to 165°C, then increased by 2°C/min to 220°C, and was subsequently held there for 17 minutes. Identification of fatty acid methyl eater peaks was based on retention time of F.A.M.E. mix C4-C24 fatty acid standard (Supelco, Bellefonte, PA, USA). Individual FA concentrations were obtained as a percent of total fatty acids. Along with individual fatty acid content, fatty acids were classified into saturated (SFA), unsaturated (UFA), monounsaturated (MUFA), polyunsaturated (PUFA), short-chain (4 to 10 carbons), medium-chain (12 to 16 carbons), or long-chain (17 to 22 carbons) fatty acid groups.
Data

In the final dataset, fatty acid analysis was completed on N=1,958 milk samples from 372 cows (average 5.26 samples per cow; range 1 to 13 samples per cow). The large range in the number of samples per cow was due to the frequency of tests for a herd, the condition of the sample at the time of analysis, and the condition of the cow (entering their dry period, being in subnormal health, or leaving the herd). Individual and groups of fatty acids were converted from g/100g of fatty acid to g/dL of milk using the fat content determined during milk recording by MIR spectroscopy and an approximation of the density of milk. Total fat content was only obtainable for 1,852 of the samples and thus fewer samples had fatty acids expressed as g/dL of milk. To maintain as much variation as possible, but still remove extreme values, individual fatty acid amounts (g/100g) more than 5 standard deviations away from the mean were removed. Due to the nature of the fatty acid determination, the entire record was deleted if one value was deemed an outlier. After editing, there were 1,919 samples with fatty acids expressed as g/100g and 1,813 samples expressed as g/dL. The number of cows, herds and per-cow samples after editing by breed is shown in Table 3.1.

MIR Prediction Calibration Sets

Four measures of fatty acid content were evaluated for building fatty acid prediction equations from MIR spectra: fatty acid content measured (1) as g/100g of fatty acid; (2) as g/dL of milk; (3) as the natural logarithm of the g/dL fatty acid content plus a constant of one (in order to include samples with values of zero); and (4) by eliminating samples with fatty acid content coming from overrepresented areas of the fatty acid content (in ln(g/dL+1)) distribution. The natural logarithm of fatty acid contents on a milk basis was considered in order to normalize the distribution of fatty acid contents, which was positively skewed for some fatty acids. The final method of eliminating samples was examined because histograms of the fatty acid concentrations showed that there were many more samples closer to the mean concentrations than there were in the tails of the distributions (a leptokurtic distribution). For the purpose of building prediction equations, having equal numbers of samples across the full range of fatty acid concentrations would be ideal, but impractical to collect. Calibrations developed with sample sets having a Gaussian distribution may cause predictions of future samples to regress toward the mean, a phenomenon known as the “Dunne” effect (Dunne and Anderson, 1976; Williams, 2001). This
effect will be more pronounced in sample sets with very large variance and low correlation between the infrared spectrum and reference values (Williams, 2007). Thus, a more uniform distribution was generated from the original calibration samples using a uniform random selection procedure. To create the subset independently for each individual fatty acid or each fatty acid group, the natural log transformed g/dL fatty acid records were partitioned into bins equal to one one-hundredth of the range of that particular fatty acid. A maximum of 18 samples per bin were randomly selected from each bin (1% of the total number of samples). So, if a bin had only 18 or fewer samples, then all of the records in that particular bin were included in the subset. The number of samples used in the training set to create the prediction equation using the subsets was therefore far fewer than the other approaches and varied between the different fatty acids depending on the original distribution. This subset selection process was repeated five times to create five different subsets for each individual fatty acid or group. An examination of the distribution of all created calibration sets was performed using the UNIVARIATE procedure in SAS (SAS Institute, 2013).

**MIR Prediction Models**

All prediction equations were constructed by partial least squares regression using the PLS procedure of SAS (SAS Institute, 2013). First, all samples with both MFG size and spectral data available were included in the calibration set and the MFG was regressed on the spectral data using PLS. The root mean square error for standardized predictors was examined for each milk sample as a measure of the distance between the data point and the model plane in X-space. Sample spectra with a mean square error greater than 3 standard deviations above the mean value were considered outliers and omitted from the analysis. The PLS procedure with leave-one-out cross-validation was then used on the remaining data to produce the final calibration equation. In particular, one milk sample at a time was reserved as holdout data for testing and all other samples were used to train the model. This process was repeated until each sample has been predicted in turn, with the validation errors saved each time and then averaged to create the standard error of cross-validation (SECV). The coefficient of determination of cross-validation ($R^2_C$), which indicates the proportion of the sample variation explained by the regression model, was used to assess how well the prediction fit the data. Additionally, the ratio of performance to deviation (RPD), calculated as the ratio of the standard deviation of the calibration set to the
SECV, was determined as an additional measure of model utility. For RPD a higher value is
desired and models with an RPD greater than two are said to produce good predictions (De
Marchi et al., 2011). Predictions were created for each of the five subsets of fatty acids and the
final fitting statistics were averaged across the five repeats.

Soyeurt et al. (2011) suggested that prediction equations with $R^2_{CV} > 0.95$ could be used
for payment purposes, and equations with an $R^2_{CV} > 0.75$ could be used for animal breeding
purposes. However, Cecchinato et al. (2009) showed that despite low calibration $R^2$ for their
MIR predicted milk coagulation properties, the genetic correlation between the measured and
predicted values were large and predicted values could be used successfully as indicator traits to
genetically improve milk coagulation properties.

RESULTS AND DISCUSSION

Descriptive Statistics

Descriptive statistics of the gas chromatography-measured individual and grouped fatty
acids expressed in g/100g and g/dL are summarized in Table 3.2. The fatty acids appearing in the
highest concentrations were C16:0, C18:1n9c, and C14:0. Sufficient variation was observed in
the measured fatty acids to produce calibration equations, which is a signature of sampling from
a wide range of management practices. The coefficient of variation (CV) for milk samples in the
full dataset expressed as g/100g ranged from 6.87 to 213.55%, which is comparable, although on
average slightly lower, to that of Soyeurt et al. (2011). Not surprisingly, the CV for most fatty
acids increased when the subset method was used to represent fatty acid content. Further, fatty
acids given as g/dL showed slightly more variation than that given as g/100g due to differences
in the fat content of the milk samples. Fatty acids that were on average present in very small
quantities had large CV values. This trend was particularly true for C22:6n3, which had the
lowest concentration of all measured fatty acids. Approximately 78% of the milk samples had
recorded concentrations of zero for this fatty acid, which could in part be due to minimum
detection values. Consequently, C22:6n3 had a highly positively skewed, leptokurtic distribution
and the associated mean and standard deviation were greatly affected by the presence of zeros.
However, upon omission of almost all recorded zeros, the CV decreased to a value similar to that
of the other fatty acids.
**Prediction Equations**

The fitting statistics for all of the prediction models are shown in Table 3.3. The range in $R^2_{cv}$ of the predictions of fatty acids ranged from 0.09 to 0.76 for g/100g of fatty acid and 0.07 to 0.93 for g/dL of milk. For all but C18:2n6t, which had an $R^2_{cv}$ close to zero, the $R^2_{cv}$ value increased when fatty acids were expressed as g/dL milk compared to g/100g fatty acid. The only individual fatty acid to achieve an $R^2_{cv}$ of 0.70 when expressed as g/100g was C18:1n9c. The fatty acid groups of saturated, monounsaturated, unsaturated, medium-chain, and long-chain also had $R^2_{cv}$ values greater than 0.70. When prediction models were created for fatty acids as g/dL, an additional five individual fatty acids had $R^2_{cv}$ values over 0.70. These findings are in line with Soyeurt et al. (2006a) and Rutten et al. (2009), who also observed higher accuracies when fatty acids were expressed on a per milk basis versus per fat basis. This is explained partly by the fact that milk samples contain different amounts of total fat and samples with the same relative concentrations of fatty acids can contain very different total quantities of the fatty acids. The MIR spectrum correlates to a greater extent with the total amount of a fatty acid than the proportion of fatty acids.

In most cases, the individual or groups of fatty acids examined in the present study that appeared in greater concentrations had the highest $R^2_{cv}$. The eleven individual or grouped fatty acids that were most prevalent in the data, were the only ones for which $R^2_{cv} > 0.70$ when expressed as g/dL. As well, the fatty acids appearing in negligible amounts did not predict well enough to be useful. The relationship between the fatty acid concentration and predictive model performance has also been identified and discussed by Soyeurt et al. (2006a) and De Marchi et al. (2011). Rutten et al. (2009) modeled the relationship between fatty acid concentration (g/dL) and prediction $R^2$ and reported an $R^2$ value of 0.64.

Compared to De Marchi et al. (2011), the present study had greater prediction accuracies for most of the fatty acids examined by both, but the former results were based on a smaller number of milk samples ($N = 267$) and only Brown Swiss cows. The one exception was C8:0, which predicted poorly in our samples. Soyeurt et al. (2011) used a diverse fatty acid dataset of 517 samples, and developed prediction equations with larger $R^2_{cv}$ than reported here for examined fatty acids, with models performing marginally or considerably better. With a larger number of samples ($N = 3,622$) studied by Rutten et al. (2009), validation $R^2$ reported for fatty acids predictions were also greater than observed presently. Most notably the short-chain fatty
acids in our data were predicted unsatisfactorily, while Rutten et al. (2009) observed validation $R^2$ values greater than 0.90 for all C4:0, C6:0, and C8:0. Ferrand et al. (2011) also showed these short-chain fatty acids could predict well.

The better results in other studies could be a result of differences in the variability in the calibration data, and the statistical procedures used (De Marchi et al., 2014). The procedures used for gas chromatography measurement of fatty acids could affect the end accuracy of the predictions. Also, importantly, the calibration set needs to incorporate all of the variation expected to be in the population to be predicted. The methodologies for developing the prediction equations and the different pre-treatments other studies have tried on the received spectra could also create differences. Improved accuracies of prediction equations have been achieved by other studies by using first-derivative pre-processing of the spectra or wavelength selection before PLS regression (Soyeurt et al., 2011; Ferrand-Calmels et al., 2014).

**Logarithmic Transformation Calibration Set**

Partial least squared regression methods perform best with symmetrically distributed data. The distribution of the measured fatty acids used in the calibration are generally not symmetrical as shown by the skewness values reported in Table 3.4. In most cases, the skewness of the distribution increased when fatty acids were converted to g/dL from g/100g. Just over half of the examined individual and groups of fatty acids had skewness greater than one when expressed as g/dL. After a natural logarithmic transformation was performed to produce the calibration dataset, the data became more symmetrical although seven fatty acids still had skewness greater than one.

However, the performance of the prediction models based on the log-transformed calibration sets did not greatly affect the $R^2_{cv}$ and RPD values (Table 3.3), with the exception of that for C6:0 and C8:0. The log transformation on these two fatty acids improved the model $R^2_{cv}$, although these values still remained low. The fatty acids with the most skewed distributions are those detected in small quantities in milk and did not have adequate prediction to start and thus, no noticeable improvements were noted.
**Sample Subset Calibration Set**

For most fatty acids, the created subsets succeeded in creating more uniform distributions suitable for calibration. The amount of excess kurtosis of the calibration set distributions decreased with the log transformation for all fatty acids but C18:2n6t using the subset samples (Table 3.4). The distribution of C18:2n6t exhibited the highest skewness and kurtosis of all the fatty acids measured. The $R^2_{cv}$ of the predictive models improved when using the subset data for all but five fatty acids (Table 3.3). These fatty acids were C6:0, C8:0, C11:0, C13:0, and C22:6n3, all of which had $R^2_{cv}$ values less than 0.5 and thus seemed unsatisfactory for prediction.

The subset calibration set models were repeated five times each with a different randomly selected subset. The $R^2_{cv}$ achieved by the five repeats were not identical and the differences between them varied depending on the component. C22:6n3 exhibited the most dramatic differences in $R^2_{cv}$ between the five repeats with a range of 0.21. This is likely due to the much smaller sample size of 392 milk samples used, a resultant of the very large number of samples having no quantifiable concentration. However, the performance of the prediction of this fatty acid is far below the level of being useful for all tested calibration sets and this large $R^2_{cv}$ range is not indicative of problems for other fatty acids. Rutten et al. (2010) examined the relationship between the number of samples used for calibration and the validation $R^2$ in MIR predicted fatty acids. When using a small number of samples (N = 100) they observed a large range in $R^2$ values from 0.05 to 0.30 for C16:0, but as the number of samples increased, the magnitude of the observed range decreased. Apart from C22:6n3, the largest range in $R^2_{cv}$ values observed in the present study was 0.09 and the average was 0.03. In general, it is expected that increasing the number of samples within a calibration set will produce better predictions and more robust models. Rutten et al. (2010) also observed an improvement in their $R^2$ with increasing sample numbers. The current study largely saw an increase in $R^2_{cv}$ when the number of samples was decreased. The reduced sample numbers in the calibration sets were still sufficiently large for the most part but these results illustrate the importance of which samples are included in the calibration set. Note that the number of unique or influential samples could be more important than the total number of samples. However, external validation will be required to satisfactorily identify the precision of the present models and how well they perform on another population of milk samples.
An alternate method for selecting subsets that take into consideration the sample spectra itself should be examined in the future. Assuming that samples with like MIR spectra are compositionally similar, perhaps samples with near identical spectra can be removed from the calibration set. This selection process would lessen the Gaussian distribution of calibration samples and largely eliminate the Dunne effect, which is why this method is attractive for selecting samples for analytic analysis. However, due to the complex nature of the composition of milk, it may still be challenging to uncover samples differing in a minor milk component of interest if it does not dominate the spectra. Additionally, specific milk components may have different ideal calibration sets. By randomly selecting samples out of a group with near identical quantities of one fatty acid, inadvertently, samples with similar composition for another, possibly correlated milk trait such as total fat content may be selected together. This could cause the predictive models to inappropriately put strength on the regions of the spectrum relating to the other component and incorporate the correlation in the model. When measured milk components are readily available, along with spectral data, coupling sample selection strategies involving both sample composition and spectral information could be an improvement upon randomly selecting samples with similar composition to produce the more uniform distribution. Such selection strategies may aid in ensuring variability in regards to other milk components within a group of samples with the same quantity of the fatty acid of interest. As a substitute to spectra, selecting samples based on the trait of interest while also considering other known, measured milk component traits influencing the spectrum could also be investigated.

**CONCLUSIONS**

The use of MIR spectra to predict fatty acid content in bovine milk was examined for the Canadian dairy population. The accuracy of the predictions depended on the fatty acid examined and the calibration set used to create the equation. The greatest $R^2_{cv}$ was achieved for fatty acids with high concentrations in milk and when they were expressed on a per milk volume basis. Excluding excess samples from overabundant regions of the distribution made further improvements to the equations and can be further investigated. The predictions for some of the fatty acids are sufficient for monitoring changes in fatty acid profiles and for use in animal breeding programs for potential genetic changes. Predicted fatty acids from equations with lower $R^2_{cv}$ may still be useful as indicators for actual fatty acid contents. Future research will further
examine the ideal calibration set for different fatty acids and spectral pre-treatment procedures to improve prediction equations as well as their utility in genetic improvement programs.
Table 3.1. The number of cows, herds, and samples after editing with fatty acid content expressed on a fat (g/100g) and milk (g/dL) basis by breed of cow.

<table>
<thead>
<tr>
<th></th>
<th>In Fat (g/100g)</th>
<th>In Milk (g/dL)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cows</td>
<td>Herds</td>
</tr>
<tr>
<td>Ayrshire</td>
<td>59</td>
<td>7</td>
</tr>
<tr>
<td>Brown Swiss</td>
<td>25</td>
<td>3</td>
</tr>
<tr>
<td>Holstein</td>
<td>228</td>
<td>29</td>
</tr>
<tr>
<td>Jersey</td>
<td>58</td>
<td>8</td>
</tr>
<tr>
<td>Total</td>
<td>370</td>
<td>44</td>
</tr>
</tbody>
</table>

1 Three herds had multiple breeds
Table 3.2. Descriptive statistics of gas chromatography determined fatty acid content on a fat (g/100g; N = 1,919) and milk (g/dL; N = 1,813) basis.

<table>
<thead>
<tr>
<th>Fatty Acid</th>
<th>In Fat (g/100g)</th>
<th>CV</th>
<th>In Milk (g/dL)</th>
<th>Mean</th>
<th>CV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual Fatty Acids</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C4:0</td>
<td>3.759</td>
<td>18.946</td>
<td>0.163</td>
<td>26.152</td>
<td></td>
</tr>
<tr>
<td>C6:0</td>
<td>1.877</td>
<td>34.955</td>
<td>0.082</td>
<td>41.102</td>
<td></td>
</tr>
<tr>
<td>C8:0</td>
<td>1.644</td>
<td>36.822</td>
<td>0.071</td>
<td>40.695</td>
<td></td>
</tr>
<tr>
<td>C10:0</td>
<td>3.513</td>
<td>27.495</td>
<td>0.153</td>
<td>35.102</td>
<td></td>
</tr>
<tr>
<td>C11:0</td>
<td>0.444</td>
<td>51.559</td>
<td>0.019</td>
<td>54.543</td>
<td></td>
</tr>
<tr>
<td>C12:0</td>
<td>3.816</td>
<td>29.294</td>
<td>0.167</td>
<td>37.798</td>
<td></td>
</tr>
<tr>
<td>C13:0</td>
<td>0.177</td>
<td>82.193</td>
<td>0.008</td>
<td>85.646</td>
<td></td>
</tr>
<tr>
<td>C14:0</td>
<td>12.356</td>
<td>17.413</td>
<td>0.536</td>
<td>26.065</td>
<td></td>
</tr>
<tr>
<td>C14:1</td>
<td>1.115</td>
<td>33.222</td>
<td>0.049</td>
<td>38.857</td>
<td></td>
</tr>
<tr>
<td>C15:0</td>
<td>1.224</td>
<td>23.832</td>
<td>0.053</td>
<td>31.199</td>
<td></td>
</tr>
<tr>
<td>C16:0</td>
<td>31.119</td>
<td>13.121</td>
<td>1.355</td>
<td>24.034</td>
<td></td>
</tr>
<tr>
<td>C16:1</td>
<td>1.938</td>
<td>23.523</td>
<td>0.084</td>
<td>31.087</td>
<td></td>
</tr>
<tr>
<td>C17:0</td>
<td>0.723</td>
<td>23.322</td>
<td>0.031</td>
<td>31.432</td>
<td></td>
</tr>
<tr>
<td>C17:1</td>
<td>0.204</td>
<td>57.332</td>
<td>0.009</td>
<td>62.881</td>
<td></td>
</tr>
<tr>
<td>C18:0</td>
<td>9.775</td>
<td>24.343</td>
<td>0.421</td>
<td>30.548</td>
<td></td>
</tr>
<tr>
<td>C18:1n9t</td>
<td>2.224</td>
<td>38.665</td>
<td>0.095</td>
<td>40.353</td>
<td></td>
</tr>
<tr>
<td>C18:1n9c</td>
<td>18.711</td>
<td>21.955</td>
<td>0.807</td>
<td>28.057</td>
<td></td>
</tr>
<tr>
<td>C18:2n6t</td>
<td>0.218</td>
<td>58.874</td>
<td>0.009</td>
<td>59.054</td>
<td></td>
</tr>
<tr>
<td>C18:2n6c</td>
<td>1.929</td>
<td>31.730</td>
<td>0.083</td>
<td>34.635</td>
<td></td>
</tr>
<tr>
<td>C18:3n3</td>
<td>0.726</td>
<td>44.995</td>
<td>0.031</td>
<td>47.571</td>
<td></td>
</tr>
<tr>
<td>C18:2n9c,12c</td>
<td>0.625</td>
<td>43.882</td>
<td>0.027</td>
<td>45.626</td>
<td></td>
</tr>
<tr>
<td>C22:6n3</td>
<td>0.047</td>
<td>213.545</td>
<td>0.002</td>
<td>215.933</td>
<td></td>
</tr>
</tbody>
</table>

Fatty Acid Groups\(^1\)

| Saturated               | 70.427          | 6.868   | 3.060   | 20.666 |
| Monounsaturated         | 24.193          | 17.948  | 1.043   | 24.750 |
| Polyunsaturated         | 3.546           | 23.159  | 0.152   | 26.762 |
| Unsaturated             | 27.739          | 17.439  | 1.196   | 24.042 |
| Short-chain             | 10.790          | 16.586  | 0.468   | 25.450 |
| Medium-chain            | 52.189          | 11.640  | 2.272   | 23.066 |
| Long-chain              | 35.184          | 18.597  | 1.515   | 25.231 |

\(^1\)Short-chain (4 to 10 carbons), medium-chain (12 to 16 carbons), and long-chain (17 to 22 carbons) fatty acid groups
Table 3.3. Fitting statistics of each calibration equation estimating fatty acid concentrations using the calibration datasets expressed as g/100g fat (F), g/dL milk (M), ln(g/dL+1) (LN), and the subsets (S)\(^1\). Bold face represents the model with the highest value.

<table>
<thead>
<tr>
<th>Fatty Acid</th>
<th>N samples</th>
<th>(R_{cv}^2)</th>
<th>RPD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F</td>
<td>M</td>
<td>LN</td>
</tr>
<tr>
<td>Individual Fatty Acids</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C4:0</td>
<td>1,876</td>
<td>1,773</td>
<td>1,773</td>
</tr>
<tr>
<td>C6:0</td>
<td>1,879</td>
<td>1,777</td>
<td>1,777</td>
</tr>
<tr>
<td>C8:0</td>
<td>1,877</td>
<td>1,773</td>
<td>1,768</td>
</tr>
<tr>
<td>C10:0</td>
<td>1,881</td>
<td>1,773</td>
<td>1,778</td>
</tr>
<tr>
<td>C11:0</td>
<td>1,873</td>
<td>1,772</td>
<td>1,772</td>
</tr>
<tr>
<td>C12:0</td>
<td>1,873</td>
<td>1,775</td>
<td>1,774</td>
</tr>
<tr>
<td>C13:0</td>
<td>1,873</td>
<td>1,786</td>
<td>1,786</td>
</tr>
<tr>
<td>C14:0</td>
<td>1,868</td>
<td>1,763</td>
<td>1,764</td>
</tr>
<tr>
<td>C14:1</td>
<td>1,881</td>
<td>1,779</td>
<td>1,779</td>
</tr>
<tr>
<td>C15:0</td>
<td>1,881</td>
<td>1,778</td>
<td>1,779</td>
</tr>
<tr>
<td>C16:0</td>
<td>1,865</td>
<td>1,779</td>
<td>1,779</td>
</tr>
<tr>
<td>C16:1</td>
<td>1,876</td>
<td>1,779</td>
<td>1,779</td>
</tr>
<tr>
<td>C17:0</td>
<td>1,890</td>
<td>1,764</td>
<td>1,770</td>
</tr>
<tr>
<td>C17:1</td>
<td>1,876</td>
<td>1,774</td>
<td>1,774</td>
</tr>
<tr>
<td>C18:0</td>
<td>1,880</td>
<td>1,763</td>
<td>1,763</td>
</tr>
<tr>
<td>C18:1n9t</td>
<td>1,881</td>
<td>1,779</td>
<td>1,779</td>
</tr>
<tr>
<td>C18:1n9c</td>
<td>1,871</td>
<td>1,777</td>
<td>1,777</td>
</tr>
<tr>
<td>C18:2n6t</td>
<td>1,876</td>
<td>1,774</td>
<td>1,773</td>
</tr>
<tr>
<td>C18:2n6c</td>
<td>1,881</td>
<td>1,778</td>
<td>1,778</td>
</tr>
<tr>
<td>C18:3n3</td>
<td>1,878</td>
<td>1,775</td>
<td>1,775</td>
</tr>
<tr>
<td>C18:2n9c,12c</td>
<td>1,881</td>
<td>1,778</td>
<td>1,778</td>
</tr>
<tr>
<td>C22:6n3</td>
<td>1,874</td>
<td>1,773</td>
<td>1,773</td>
</tr>
<tr>
<td>Fatty Acid Groups(^2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Saturated</td>
<td>1,879</td>
<td>1,771</td>
<td>1,778</td>
</tr>
<tr>
<td>Monounsaturated</td>
<td>1,875</td>
<td>1,777</td>
<td>1,776</td>
</tr>
<tr>
<td>Polyunsaturated</td>
<td>1,880</td>
<td>1,770</td>
<td>1,770</td>
</tr>
<tr>
<td>Unsaturated</td>
<td>1,881</td>
<td>1,777</td>
<td>1,778</td>
</tr>
<tr>
<td>Short-chain</td>
<td>1,870</td>
<td>1,764</td>
<td>1,763</td>
</tr>
<tr>
<td>Medium-chain</td>
<td>1,868</td>
<td>1,778</td>
<td>1,778</td>
</tr>
<tr>
<td>Long-chain</td>
<td>1,874</td>
<td>1,771</td>
<td>1,767</td>
</tr>
</tbody>
</table>

\(^1\)\(R_{cv}^2\) - Coefficient of determination of cross validation; RPD - Ratio of performance deviation.

\(^2\)Short-chain (4 to 10 carbons), medium-chain (12 to 16 carbons), and long-chain (17 to 22 carbons) fatty acid groups.
Table 3.4. Measured skewness and excess kurtosis of the distribution of the calibration samples for fatty acids expressed as g/100g fat (F), g/dL milk (M), ln(g/dL+1) (LN), and the subsets (S).

<table>
<thead>
<tr>
<th>Fatty Acid</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F</td>
<td>M</td>
</tr>
<tr>
<td>Individual Fatty Acids</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C4:0</td>
<td>0.484</td>
<td>0.948</td>
</tr>
<tr>
<td>C6:0</td>
<td>0.429</td>
<td>0.773</td>
</tr>
<tr>
<td>C8:0</td>
<td>1.190</td>
<td>1.083</td>
</tr>
<tr>
<td>C10:0</td>
<td>0.611</td>
<td>1.094</td>
</tr>
<tr>
<td>C11:0</td>
<td>1.041</td>
<td>1.293</td>
</tr>
<tr>
<td>C12:0</td>
<td>0.453</td>
<td>1.019</td>
</tr>
<tr>
<td>C14:0</td>
<td>-0.177</td>
<td>0.569</td>
</tr>
<tr>
<td>C14:1</td>
<td>0.265</td>
<td>0.665</td>
</tr>
<tr>
<td>C15:0</td>
<td>0.413</td>
<td>0.856</td>
</tr>
<tr>
<td>C16:0</td>
<td>0.102</td>
<td>0.624</td>
</tr>
<tr>
<td>C16:1</td>
<td>0.756</td>
<td>1.097</td>
</tr>
<tr>
<td>C17:0</td>
<td>0.073</td>
<td>0.574</td>
</tr>
<tr>
<td>C17:1</td>
<td>0.217</td>
<td>0.771</td>
</tr>
<tr>
<td>C18:0</td>
<td>0.352</td>
<td>0.756</td>
</tr>
<tr>
<td>C18:1n9t</td>
<td>1.109</td>
<td>1.158</td>
</tr>
<tr>
<td>C18:1n9c</td>
<td>0.735</td>
<td>1.602</td>
</tr>
<tr>
<td>C18:2n6c</td>
<td>1.026</td>
<td>1.389</td>
</tr>
<tr>
<td>C18:3n3</td>
<td>0.779</td>
<td>0.773</td>
</tr>
<tr>
<td>C18:2n9c,12c</td>
<td>1.173</td>
<td>1.152</td>
</tr>
<tr>
<td>Fatty Acid Groups</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Saturated</td>
<td>-0.493</td>
<td>0.556</td>
</tr>
<tr>
<td>Monounsaturated</td>
<td>0.670</td>
<td>1.476</td>
</tr>
<tr>
<td>Polyunsaturated</td>
<td>0.317</td>
<td>0.532</td>
</tr>
<tr>
<td>Unsaturated</td>
<td>0.530</td>
<td>1.315</td>
</tr>
<tr>
<td>Short-chain</td>
<td>0.629</td>
<td>0.857</td>
</tr>
<tr>
<td>Medium-chain</td>
<td>-0.392</td>
<td>0.520</td>
</tr>
<tr>
<td>Long-chain</td>
<td>0.387</td>
<td>1.156</td>
</tr>
</tbody>
</table>

1Short-chain (4 to 10 carbons), medium-chain (12 to 16 carbons), and long-chain (17 to 22 carbons) fatty acid groups.
Chapter 4
GENETIC CORRELATIONS OF MID-INFRARED PREDICTED MILK FATTY ACID GROUPS WITH MILK PRODUCTION TRAITS

ABSTRACT
The objective of this research was to estimate the genetic correlations between milk MIR predicted fatty acid groups and production traits in first-parity Canadian Holsteins. Contents of short-chain, medium-chain, long-chain, saturated, and unsaturated fatty acid groupings in milk samples can be predicted using mid-infrared spectral data for cows enrolled in milk recording programs. Predicted fatty acid group contents were obtained for 49,127 test-day milk samples from 10,029 first-parity Holstein cows in 810 Canadian herds. Milk yield, fat and protein yield, fat and protein percentage, fat:protein ratio, and SCS were also available for these test-days. Genetic parameters were estimated for the fatty acid groups and production traits using multiple-trait random regression test-day models by Bayesian methods via Gibbs sampling. Regression curves were modeled using Legendre polynomials of order 5 for the random effects. A total of 3 separate 8- or 9-trait analyses were performed including the 5 fatty acid groups with different combinations of the production traits. Average daily genetic correlations were negative and similar to each other for the fatty acid groups with milk yield (-0.560 to -0.494) and with protein yield (-0.263 to -0.192). Weak and positive average daily genetic correlations were found between SCS and the fatty acid groups (from 0.237 to 0.274). Stronger genetic correlations with fat yield, fat and protein percentage, and fat:protein ratio were found with medium-chain and saturated fatty acid groups compared to those with long-chain and unsaturated fatty acid groups. Genetic correlations were very strong between the fatty acid groups and fat percentage, ranging between 0.814 for unsaturated and 0.973 for saturated fatty acids. Daily genetic correlations from 5 to 305 days in milk with milk, protein, and SCS traits showed similar patterns for all fatty acid groups. The daily genetic correlation with fat yield at the beginning of lactation were decreasing for long-chain and unsaturated fatty acid groups and increasing for short- and medium-chain and saturated fatty acids. Genetic correlations between fat percentage and fatty acids were low but increasing at the beginning of lactation for short- and medium-chain and saturated fatty acids, but starting higher and remaining more constant for long-chain and unsaturated fatty acid groups. These results can be used in defining fatty acid traits and breeding objectives.
**Key words:** fatty acid, production trait, genetic correlation

## INTRODUCTION

There is growing interest in broadening selection objectives in dairy cattle breeding to include novel traits to improve milk quality, and cow health and fertility so as to complement selection for production traits. The composition of milk fat is meaningful for human health and nutrition, but also as a potential indicator for the metabolic and health status of the cow. Emerging consumer concern over the health effects of the products they consume has elicited research into the possibility of altering the fatty acid content of milk. Milk fat is rich in saturated fatty acids, and consumption of some of these saturated fatty acids may contribute to an increased risk of heart diseases, atherosclerosis, and weight gain (Haug, 2007). However, bovine milk fat is also a source of many fatty acids with potential health benefits (Parodi, 2004). A reduction in the proportion of saturated fatty acids in bovine milk may therefore be of great interest, along with increased proportions of those fatty acids promoting good health.

The fatty acid profile of milk can be altered through the feeding of specialized diets to lactating dairy cows (Palmquist et al., 1993; Chilliard et al., 2000; Kliem and Shingfield, 2016). However, these changes are only temporary and depend on feed availability and prices. As well, there are limitations to the amount of fat that can be fed to dairy cows and the availability of these dietary lipids to the mammary gland to be used for milk production (Kliem and Shingfield, 2016). The fatty acid composition of milk is also influenced by the genetics of the cow (Bobe et al., 2008; Soyeurt et al., 2008; Stoop et al., 2008; Garnsworthy et al., 2010; Bastin et al., 2011; Pegolo et al., 2016), thus providing the possibility for genetic selection to permanently enhance the milk fatty acid profile for the entire cow population.

The implementation of a new trait into a selection program requires an affordable means of phenotyping the trait routinely, on a large scale. Mid-infrared (MIR) spectroscopy is already used to record fat and protein percentage, milk urea nitrogen, and beta-hydroxybutyrate contents in milk during routine milk recording. More recently, studies have shown that milk fatty acid contents can also be predicted from the MIR spectra of a milk sample (Soyeurt et al., 2006; Rutten et al., 2009; De Marchi et al., 2011; Soyeurt et al., 2011; Ferrand-Calmels et al., 2014; Fleming et al., 2016a). The prediction of milk fatty acid contents using MIR technology facilitates the collection of large numbers of these phenotypes for all milk-recorded cows.
Genetic evaluation and selection for these traits is thus achievable for the dairy industry. In order to avoid unwanted correlated responses with other important traits, knowledge of these genetic correlations is important prior to implementation of fatty acid composition traits into breeding goals. Few studies have looked to estimate the daily genetic correlations of these traits across the lactation. The objective of this study was therefore to estimate the genetic correlations of milk fatty acid group contents with milk production traits in the Canadian Holstein population using multiple trait animal models.

**MATERIALS AND METHODS**

**Data**

Milk MIR spectra obtained during routine milk recording from one of two MilkoScan FT6000 spectrometers (FOSS, Hillerød, Denmark) at CanWest DHI (Guelph, ON) or Valacta (Sainte-Anne-de-Bellevue, QC) milk laboratories were received and stored in a database. Spectra were collected between January 2013 and July 2015 and standardization of spectra between the two machines was performed per Bonfatti et al. (2016). The database contained a total of 2,053,396 spectra from Holstein cow milk samples.

Prediction equations developed by Fleming et al. (2016a) for five groups of fatty acids expressed as the natural logarithm of grams per deciliter of milk were applied to the historical spectra. Fatty acid groups were defined by saturation (saturated (SFA) and unsaturated (UFA)) and by chain-length (short-chain (SCFA; 4 to 10 carbons), medium-chain (MCFA; 12 to 16 carbons), and long-chain (LCFA; 17 to 22 carbons)). The individual fatty acids used to devise the groups are given in Table 4.1. Fatty acid contents were examined on a per milk basis instead of on a per fat basis because of the greater prediction accuracies achieved. The coefficients of determination of cross-validation ($R^2_C$) for the prediction equations were 0.93, 0.83, 0.73, 0.89, and 0.80 for SFA, UFA, SCFA, MCFA, and LCFA groups, respectively. Fatty acid predictions were deleted for samples with MIR spectral data considered as outliers and dissimilar to those used to develop the calibrations. Spectral outliers were determined using the root mean square error for standardized predictors, calculated when the prediction equations by Fleming et al. (2016a) were applied. Sample spectra with a mean square error greater than 3 standard deviations above the mean value were considered outliers and removed. All five groups of fatty acids were predicted for 1,957,353 Holstein milk samples with spectra saved in the database.
Test-day production records included milk, fat, and protein yields, somatic cell count (SCC), and fat to protein ratio (F:P), and were obtained from Canadian Dairy Network (Guelph, ON, Canada) for the test-days with fatty acid contents predicted. The SCC was log-transformed to somatic cell score (SCS) according to the formula proposed by Ali and Shook (1980). The total number of milk samples with predicted FA contents and test-day records was 1,780,089, from 514,795 Holstein cows in 6,768 herds.

The imposed data restrictions and edits for the genetic analysis were the same as those defined by Narayana et al. (2016). In brief, only test-day records between 5 and 305 days in milk of first-parity cows, with calving between 19 and 43 months of age were considered. Included cows were required to be from herds with at least 70 cows and to have had their first test-day record within the first 50 days of lactation. A minimum of 4 records per herd on a test-day was required. Strict data restrictions were required due to the low density of test-day records with fatty acid records in the full data set. Only a small proportion (approximately 10%) of milk samples collected during milk recording had their MIR spectra outputted because only two milk spectrometers in Canadian DHI currently have these capabilities. The process by which milk samples were selected to go through these lines and have spectra collected was mostly random. The final edited dataset consisted of 49,127 test day records from 10,029 first-parity Holstein cows from 810 herds. The number of test-days per cow ranged from 4 to 10 and averaged approximately 4.9 test-days per cow. A full pedigree file, going back as many generations as available, containing 76,074 individuals was provided by the Canadian Dairy Network.

**Model**

Three separate 8- or 9-trait random regression test-day models were used for the genetic analysis. The 3 analyses were performed for the following combinations of traits: 1) 5 FA groups, milk, fat and protein yield, SCS; 2) 5 FA groups, milk yield, fat and protein percentage, SCS; and 3) 5 FA groups, milk yield, F:P, SCS. The model used for the genetic analyses was developed and described by Narayana et al. (2016). The model considered for all traits can be expressed in matrix notation by the equation:

\[ y = X_c c + X_b b + Z_h h + Z_a a + Z_p p + e, \]

where \( y \) is a vector of observations; \( c \) is a vector of fixed class effects of herd-test day and days in milk (DIM; 300 classes) effects; \( b \) is a vector of fixed regression coefficients for age-season of
calving effects (4 seasons and 25 age classes); \( h \) is a vector of random regression coefficients for herd-year of calving effects; \( p \) is a vector of random regression coefficients for permanent environment effects; \( a \) is a vector of random regression coefficients for animal genetic effects; \( e \) is a vector of residuals; and \( X_c, X_h, Z_h, Z_a, \) and \( Z_p \) are incidence matrices assigning observations to effects.

Regression curves were modeled using Legendre polynomials of order 4 for the fixed herd-year effect, and using Legendre polynomials of order 5 for the random effects as defined by Jamrozik et al. (2002). Expectations and covariance structure for the random effects were given by,

\[
E(y) = X_c + X_p, E(h) = 0, E(a) = 0, E(p) = 0, E(e) = 0,
\]

and

\[
V(h) = I \otimes Q_0, V(a) = A \otimes G_0, V(p) = I \otimes P_0, V(e) = E,
\]

where \( I \) is an identity matrix; \( A \) is the additive relationship matrix; \( Q_0 \) is a covariance matrix for herd-year regression coefficients; \( G_0 \) is a (co)variance matrix of genetic regression coefficients; \( P_0 \) is a (co)variance matrix for permanent environment regression coefficients; \( E \) is a block-diagonal residual (co)variance matrix; and \( \otimes \) is the Kronecker product function. Residual (co)variances were assumed heterogeneous across 20 intervals of 15 DIM (5 to 20, 21 to 35, 36 to 50, …, 291 to 305 DIM). All random effects were assumed to be normally distributed.

Variance components were estimated by Bayesian methods via Gibbs sampling using custom-written Fortran software. Prior values were set arbitrarily to 0.01 for variances and 0 for covariances. Posterior means of (co)variance components were estimated using 120,000 samples after a burn-in of 30,000 samples for each multiple-trait model. Daily heritability was defined as the ratio of genetic variance to the sum of genetic, permanent environment, herd-year of calving, and residual variances, for each DIM from 5 to 305, and averaged across the entire lactation. Because the fatty acid groups, milk yield, and SCS traits were included in all analyses, three average daily heritabilities were calculated and then averaged to produce the overall average daily heritability. Average daily genetic correlations among traits were calculated using (co)variances of the first regression coefficients as described by Wood et al. (2003). The daily and average daily genetic correlations between the FA groups and milk yield and SCS were averaged across the three analyses.

55
RESULTS AND DISCUSSION

Phenotypic Description

Descriptive statistics of the fatty acid groups and production traits of the edited data set are given in Table 4.2. The raw phenotypic correlations between the fatty acid groups and the production traits are given in Table 4.3. Phenotypically, the fatty acid contents were most strongly correlated with fat percentage. The fatty acid contents are expressed on a per milk basis and are therefore most similar to fat percentage, which is the amount of total milk fat in the milk by weight. In general, similar phenotypic correlations with the production traits were noted for SCFA, MCFA, and SFA and for LCFA and UFA. Most SFA are also in the SCFA or MCFA groups, while the majority of UFA are also included in the LCFA group, as shown in Table 4.1.

Heritabilities

The average daily heritabilities of the fatty acid groups and the analyzed production traits are displayed in Table 4.4. The fatty acid groups were all moderately heritable with values similar to those reported by Narayana et al. (2016) using the same dataset. On a biological basis, it was anticipated that the de novo synthesized SCFA would be more heritable than the dietary and adipose derived LCFA in milk. However, SCFA and LCFA were found to have similar heritabilities (0.229 and 0.219, respectively). This is likely a result of error introduced by the poorer prediction of the SCFA observed by Fleming et al. (2016a). The estimated heritabilities for milk, fat and protein yield, and fat and protein percentages are lower than those previously estimated for first-parity Canadian Holsteins (Miglior et al., 2007; Loker et al., 2012). The heritability for SCS estimated in the present study was 0.033, which is also lower than the 0.189 reported by Miglior et al. (2007) and the 0.15 reported by Loker et al. (2012). These discrepancies may be a function of the model used, or differences in the datasets. In the current dataset, there were a limited number of test-days available for some animals because MIR spectra was being recorded for only a limited proportion of all milk samples during DHI milk recording. For which samples spectrum data was also collected, was largely a random process. Consequently, there were fewer records per cow and the traits may not have been as well modeled across the lactation.
**Genetic Correlations**

The average genetic correlations between fatty acid groups and production traits are shown in Table 4.5. Posterior standard deviations for average daily genetic correlations ranged from < 0.001 to 0.035. For each production trait, the five fatty acid groups displayed genetic correlations of the identical sign and of similar magnitude, although differences were still present. Narayana et al. (2016) previously reported strong genetic correlations between the five groups of fatty acids (from 0.63 to 0.96). These strong genetic correlations explain the similarity of the genetic correlations observed for the different fatty acid groups with production traits.

Overall, the strongest genetic correlations were observed between the fatty acid groups and fat percentage. Of the fatty acid groups, SFA and MCFA had the strongest genetic correlation with fat percentage at 0.973 and 0.964, respectively. The genetic correlation between UFA and fat percentage was lower at 0.814. Penasa et al. (2016) also found that SFA were more strongly correlated with fat percentage than UFA (0.991 vs. 0.838). They further divided UFA into monounsaturated and polyunsaturated fatty acid groups, and found that monounsaturated fatty acids were more strongly genetically correlated with fat percentage than polyunsaturated fatty acids. The different groups of fatty acids analyzed were simply defined as proportions of the total fat in the milk so a strong genetic correlation was observed between the fatty acid groups and total milk fat percentage. The genetic correlations for fat percentage with milk yield, protein percentage, and SCS are also given in Table 4.5, for comparison purposes. The observed genetic correlations for milk yield, protein percentage, and SCS with fat percentage were similar to those found for the fatty acid groups, although somewhat stronger. Though genetic correlations were more moderate between fatty acids and milk yield, similar results were found with fat percentage, in that LCFA and SFA had weaker genetic correlations with milk yield than the other three groups did. Bastin et al. (2011) reported similar findings.

The genetic correlations for the fatty acid groups with milk yield were all negative and similar to each other, ranging from -0.494 for SCFA to -0.560 for SFA. This negative correlation is a result of the effect of dilution. These results are in line with those reported by Bastin et al. (2011, 2013) for all of the groups of fatty acids considered. Negative genetic correlations were also found between the groups of fatty acids and protein yield. Correlations were similar for all fatty acid groups but were the strongest at -0.263 between LCFA and protein yield. Bastin et al. (2011) approximated genetic correlations using EBV correlations between fatty acid groups and
protein yield and found these correlations to be negative. They also found the strongest genetic correlation out of the fatty acid groups to be with LCFA (-0.27). Genetic correlations with protein percentage were moderate to strong and went from 0.602 for LCFA to 0.717 for MCFA. The genetic correlations for protein percentage reported by Penasa et al. (2015) were lower at 0.601 and 0.566 for SFA and UFA, respectively. For groups of fatty acids, the range in genetic correlations with protein percentage from Bastin et al. (2011) were from 0.38 for LCFA to 0.60 for polyunsaturated fatty acids, again lower than the range found in the present study. Genetic correlations between fatty acid groups and F:P were again positive and moderate to strong in size. This trait is a function of fat and protein contents and thus displays similar results to those traits.

The genetic correlations with SCS were in general moderately weak and positive (0.237 to 0.270). Bastin et al. (2013) approximated daily genetic correlations between SCS and fatty acid contents from pairwise regressions among EBV and found very weak genetic association between the two. Their genetic correlations between groups of fatty acids and SCS were averaged across days in milk and the first three parities and ranged from -0.08 to 0.03. The higher genetic correlations found in the present study are likely related to the high genetic correlation between fat percentage and SCS that was also found (0.49). Miglior et al. (2007) and Loker et al. (2012) both reported negative average daily genetic correlations between fat percentage and SCS in the first parity for Canadian Holsteins (-0.161 and -0.08, respectively).

**Daily Genetic Correlations**

The genetic correlations between the fatty acid groups and production traits were not constant across the lactation. The trends for the daily genetic correlations throughout the lactation with the fatty acid groups are shown for each examined production trait in Figures 4.1 through 4.7. The course of the genetic correlations across the lactation for the fatty acid groups with any given production trait were most similar between MCFA and SFA and between LCFA and UFA. Again, this is due to strong genetic correlations with each other, and the similar makeups of the groups. The differences in the genetic correlations observed for the groups are also due to the distinct origins of the fatty acid groups. The SCFA and the majority of the MCFA are synthesized in the mammary gland de novo, while LCFA are provided to the mammary gland from blood lipids sourced from the diet or adipose tissues. Daily genetic correlations with total
fat percentage are also displayed for milk yield, protein percentage, and SCS. Fat percentage tended to follow a similar pattern as the fatty acid contents, but was found to be more strongly correlated with these production traits for the majority of the lactation, which is consistent with the average daily genetic correlations discussed above.

Figure 4.1 shows the daily genetic correlations across the lactation for milk yield with all fatty acid groups and total fat percentage. At the beginning of lactation the genetic correlations were all weak and proceeded to become more negative until approximately day 105, where values started to level off. With milk yield, all of the fatty acids displayed similar patterns for the genetic correlations as the lactation progressed, although correlations were stronger for MCFA and SFA than for the other groups. Daily genetic correlations approximated from the correlation between daily EBVs for 10 individual fatty acids, total polyunsaturated fatty acids, and fat percentage with milk yield were shown by Bastin et al. (2011) for first-parity Holsteins in Belgium. Although they did not group the fatty acids, a similar pattern to that found in the present study across the lactation was observed. They observed daily genetic correlations closest to zero at the beginning of lactation, before they became more negative for most fatty acids. This pattern was more notable for the individual long-chain fatty acids they examined. At the beginning of lactation, Bastin et al. (2011) found positive, although weak, genetic correlations between C18:1c9 and milk yield, which is similar to the genetic correlations between LCFA and UFA with milk yield in the present study. Similar results were also presented by Bastin et al. (2013) with daily genetic correlations estimated from the pair-wise regression of EBV. With this estimation, they observed the genetic correlations tending toward zero in late lactation, like were observed in our results.

Figure 4.2 shows the genetic correlations between fat yield and fatty acid groups. At the beginning of lactation, LCFA and UFA have genetic correlations of approximately 0.60 and 0.57 with fat yield, respectively. From this initial point, their genetic correlations then decrease for the majority of the lactation before increasing slightly at the end of lactation. Correlations for SCFA, MCFA, and SFA have lower correlations at the start of lactation (0.25, 0.48, and 0.55, respectively), but then show an increase in the first stage of lactation. These fatty acid groups display a maximum genetic correlation with fat yield between 55 and 105 days in milk, before weakening again. Related findings are also shown in Figure 4.3 for fat percentage with fatty acid groups. In the initial part of the lactation, correlations with SCFA, MCFA, and SFA were all low,
but increasing, until approximately 55 to 105 days in milk. In contrast, LCFA and UFA remain relatively constant during this phase of lactation. At the end of lactation, all groups of fatty acids display decreases in their genetic correlations with fat percentage. In the beginning of lactation, cows are in a state of negative energy balance and there is an increase in the mobilization of body reserves to meet the demands of lactation. As a result there are increased amounts of long-chain fatty acids, which originate from dietary and adipose lipids, that are incorporated into the milk fat (Stoop et al., 2009). Therefore, variations in the fatty acid contents at the beginning of lactation may be indicative of the physiological status of the cow. The genetic correlations observed at the first stage of lactation may reflect these changes occurring in the cow during this phase, and thus the beginning of lactation is a crucial period to examine in depth. In this stage, LCFA was found to have a greater genetic correlation with both fat yield and percentage than the other fatty acid groups did. Selection for total fat at this time point may therefore have a greater effect on LCFA contents than other fatty acids. Energy balance has implications for the health and future reproductive performance of the cow (Ingvartsen et al., 2003; Jorritsma et al., 2003; Pryce et al., 2004), and therefore breeders should be mindful of genetic correlations in early lactation.

Throughout the lactation, the trend observed for the genetic correlations with protein yield (Figure 4.4) and protein percentage (Figure 4.5) were similar for all of the fatty acid groups. For protein yield, correlations were generally weak throughout the lactation but were strongest and negative in mid to late lactation. Genetic correlations with protein percentage were much like those observed with milk yield although positive. Correlations were the strongest in mid-lactation and closer to zero at the beginning and end stages. Fat percentage genetic correlations with protein percentage showed the same trend as the fatty acid contents did.

The daily genetic correlations of F:P with FA groups are displayed in Figure 4.6. The observed trends are related to the genetic correlations between the fatty acids and total fat content. For LCFA and UFA, the genetic correlations are relatively constant through the beginning and middle of the lactation, and only weaken toward the very end. At the beginning of lactation, the genetic correlations are stronger for LCFA and UFA than those observed for SCFA, MCFA, and SFA. The genetic correlations for SCFA, MCFA, and SFA continue to increase from 5 days in milk until approximately 105 days in milk, where they hold steady. The F:P of a milk sample has been proposed as an indicator of energy status in lactating dairy cows.
(Grieve et al., 1986), as have the fatty acid contents in the milk of the cow (Palmquist et al., 1993). As well, Koeck et al. (2013) reported a genetic correlation of 0.30 between F:P and ketosis in the first 5-30 days in milk for first lactation Canadian Holsteins. Therefore, the stronger genetic correlation at the beginning of lactation between the LCFA and SFA may again relate to the physiological status of the cow in the first stage of lactation.

The fatty acid groups all showed similar patterns for genetic correlations with SCS across the lactation (Figure 4.7). Correlations were weak throughout the lactation but were the strongest at around 205 days in milk for all fatty acids, and close to zero at both the beginning and end of lactation. Total fat percentage displayed a similar trend, but genetic correlations were stronger across the lactation than those shown for the fatty acid groups. Since the genetic correlations were estimated between SCS and the fatty acid groups during all three of the multi-trait analyses, the displayed trends in daily genetic correlations are an average of these three. The daily genetic correlations found between the fatty acids and SCS estimated with the model including fat percentage were greater in magnitude than the presented average and were more similar to those found for fat percentage. Therefore, it should not be concluded that total fat percentage is more genetically correlated to SCS than fatty acid contents, but it is notable that the general trend across lactation is the same.

The inconsistencies in the genetic correlations across the lactation for fatty acids and production traits may aid in the setting of future selection goals related to fatty acid contents and be telling of how the milk fatty acid profile is impacted by the current selection for production traits. For all of the production traits, except for fat yield, genetic correlations with fatty acid groups were the strongest in mid-lactation. Fat yield, fat percentage, and F:P show different patterns in genetic correlations with the different fatty acid groups in the important early stage of lactation. Selection for total fat at this point in the lactation may increase LCFA and UFA more compared to the other fatty acid groups. This is likely from increased mobilization of body reserves and negative energy balance, and thus care needs to be given to prevent negative responses in cow health and reproduction traits. If the entire lactation is considered, then MCFA and SFA are more strongly correlated with total fat traits and selecting for increased fat may increase the proportion of these fatty acids to a greater extent. Similarly, of the 5 fatty acid groups examined, MCFA and SFA also have the strongest average daily genetic correlations with protein percentage and F:P, and therefore, these fatty acid groups could be more so affected.
by selection for protein percentage and F:P. The comparable correlations for the fatty acid
groups with milk and protein yield and SCS implies that selection for these traits will not have
major implications for milk fatty acid contents.

In the present study, the examined FA traits were expressed as a quantity in milk (g/dL).
As discussed earlier, this is the primary reason why the discovered genetic correlations were
similar to each other and similar to total milk fat percentage. In order to decrease this interaction
with fat percentage and to more effectively select for changes in the fatty acid profile, future
studies should focus on the estimation of genetic correlations between the same production traits
and fatty acids expressed on a per fat basis (g/100 g of fat). Genetic correlations between milk
fatty acids on a per fat basis and milk production traits have been previously examined by
Karijord et al. (1982), Soyeurt et al. (2007), Stoop et al. (2008), Mele et al. (2009), and Bilal et
al. (2014). Soyeurt et al. (2007) examined genetic correlations between fatty acids both on a milk
basis and on a fat basis with production traits and found that on a fat basis there were greater
differences in the genetic correlations between the different fatty acids with the production traits.
For example, the reported genetic correlations with fat percentage were 0.76 and -0.22 for
saturated and monounsaturated fatty acid groupings, respectively. Stoop et al. (2008) found
positive genetic correlations for SCFA and MCFA with fat percentage, and negative correlations
for LCFA and fat percentage. The same pattern was also noted between fatty acids and protein
percentage, and between fatty acids and fat yield. Future studies on MIR predicted fatty acids in
the Canadian dairy cattle population should expand on examining fatty acids on a per fat basis
and estimating the daily genetic correlations for these traits across the lactation. The impact of
selection for production traits on the fatty acid profile, or vice versa, may be more apparent with
fatty acids on a per fat basis.

As previously discussed, another consideration for this dataset is the data structure
present. Because of the limited number of test-days with fatty acid records available, as a result
of randomness in the process of spectral data acquisition, the determined genetic parameters may
be inappropriate for inference to the entire Canadian Holstein population. It will be imperative to
re-estimate the genetic parameters of the fatty acid groups when test-days across the whole
lactation are available with fatty acid records.
CONCLUSIONS

Fatty acid contents in milk were found to be genetically correlated with production traits already evaluated in dairy breeding programs. Genetic correlations were strongest between fatty acid groups and fat percentage, and similar in magnitude for all production traits. However, there were differences between the fatty groups genetic correlations with observed fat yield, fat and protein percentage, and F:P. Therefore, selection for these production traits could be inadvertently changing the contents of the groups of fatty acids to varying extents. Therefore, selection for these production traits could be inadvertently changing the contents of the groups of fatty acids to varying extents. Daily genetic correlations between fatty acids and production traits were not constant across the lactation. Trends in daily genetic correlations did differ for some production traits, and selection for these traits at various points in the lactation could impact milk fat composition differently. This is particularly true in early lactation for the fatty acid groups with fat yield and fat percentage traits where LCFA and UFA have stronger genetic correlations and energy balance of the cow could be a concern.
Table 4.1. Individual fatty acids included in the defined short-chain (SCFA), medium-chain (MCFA), long-chain (LCFA), saturated (SFA), and unsaturated (UFA) groups of fatty acid traits.

<table>
<thead>
<tr>
<th>Trait</th>
<th>Fatty acids included</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCFA</td>
<td>C4:0, C6:0, C8:0, C10:0</td>
</tr>
<tr>
<td>MCFA</td>
<td>C11:0, C12:0, C13:0, C14:0, C14:1, C15:0, C16:0, C16:1</td>
</tr>
<tr>
<td>LCFA</td>
<td>C17:0, C17:1, C18:0, C18:1n9t, C18:1n9c, C18:2n6t, C18:2n6c, C18:3n3, C18:2n9c12c, C22:6n3</td>
</tr>
<tr>
<td>SFA</td>
<td>C4:0, C6:0, C8:0, C10:0, C11:0, C12:0, C13:0, C14:0, C15:0, C16:0, C17:0, C18:0</td>
</tr>
<tr>
<td>UFA</td>
<td>C14:1, C16:1, C17:1, C18:1n9t, C18:1n9c, C18:2n6t, C18:2n6c, C18:3n3, C18:2n9c,12c, C22:6n3</td>
</tr>
</tbody>
</table>

Table 4.2. Descriptive statistics of test-day records in the edited dataset (N = 49,127) for short-chain (SCFA), medium-chain (MCFA), long-chain (LCFA), saturated (SFA), and unsaturated (UFA) fatty acid groups (ln(g/dL) and milk, fat, and protein yield, fat and protein percentage, fat to protein ratio (F:P) and SCS.

<table>
<thead>
<tr>
<th>Trait</th>
<th>Mean</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCFA</td>
<td>0.36</td>
<td>0.06</td>
<td>0.05</td>
<td>0.64</td>
</tr>
<tr>
<td>MCFA</td>
<td>1.14</td>
<td>0.13</td>
<td>0.50</td>
<td>1.80</td>
</tr>
<tr>
<td>LCFA</td>
<td>0.88</td>
<td>0.12</td>
<td>0.40</td>
<td>1.83</td>
</tr>
<tr>
<td>SFA</td>
<td>1.35</td>
<td>0.12</td>
<td>0.64</td>
<td>1.99</td>
</tr>
<tr>
<td>UFA</td>
<td>0.75</td>
<td>0.11</td>
<td>0.32</td>
<td>1.63</td>
</tr>
<tr>
<td>Milk (kg)</td>
<td>30.29</td>
<td>6.27</td>
<td>6.00</td>
<td>72.90</td>
</tr>
<tr>
<td>Fat (kg)</td>
<td>1.18</td>
<td>0.25</td>
<td>0.20</td>
<td>2.87</td>
</tr>
<tr>
<td>Protein (kg)</td>
<td>0.97</td>
<td>0.18</td>
<td>0.22</td>
<td>2.14</td>
</tr>
<tr>
<td>Fat %</td>
<td>3.96</td>
<td>0.63</td>
<td>0.84</td>
<td>8.00</td>
</tr>
<tr>
<td>Protein %</td>
<td>3.23</td>
<td>0.31</td>
<td>1.98</td>
<td>4.72</td>
</tr>
<tr>
<td>F:P</td>
<td>1.23</td>
<td>0.18</td>
<td>0.30</td>
<td>2.91</td>
</tr>
<tr>
<td>SCS</td>
<td>2.00</td>
<td>1.67</td>
<td>-3.64</td>
<td>9.64</td>
</tr>
</tbody>
</table>

Table 4.3. Phenotypic correlations for short-chain (SCFA), medium-chain (MCFA), long-chain (LCFA), saturated (SFA), and unsaturated (UFA) fatty acid groups and milk, fat, and protein yield, fat and protein percentage, fat to protein ratio (F:P) and SCS.

<table>
<thead>
<tr>
<th>Milk yield</th>
<th>Fat yield</th>
<th>Protein yield</th>
<th>Fat %</th>
<th>Protein %</th>
<th>F:P</th>
<th>SCS</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCFA</td>
<td>-0.28</td>
<td>0.30</td>
<td>-0.07</td>
<td>0.79</td>
<td>0.50</td>
<td>0.52</td>
</tr>
<tr>
<td>MCFA</td>
<td>-0.31</td>
<td>0.26</td>
<td>-0.06</td>
<td>0.78</td>
<td>0.58</td>
<td>0.46</td>
</tr>
<tr>
<td>LCFA</td>
<td>-0.20</td>
<td>0.28</td>
<td>-0.21</td>
<td>0.65</td>
<td>0.03</td>
<td>0.70</td>
</tr>
<tr>
<td>SFA</td>
<td>-0.33</td>
<td>0.32</td>
<td>-0.10</td>
<td>0.89</td>
<td>0.54</td>
<td>0.61</td>
</tr>
<tr>
<td>UFA</td>
<td>-0.20</td>
<td>0.27</td>
<td>-0.20</td>
<td>0.64</td>
<td>0.07</td>
<td>0.67</td>
</tr>
</tbody>
</table>

*Significant at $P < 0.0001$
Table 4.4. Average daily heritabilities ($h^2$) and posterior standard deviations (SD) of short-chain (SCFA), medium-chain (MCFA), long-chain (LCFA), saturated (SFA), and unsaturated (UFA) fatty acid groups and milk, fats and protein yield, fat and protein percentage, fat to protein ratio (F:P) and SCS.

<table>
<thead>
<tr>
<th>Trait</th>
<th>$h^2$</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCFA</td>
<td>0.229</td>
<td>0.001</td>
</tr>
<tr>
<td>MCFA</td>
<td>0.331</td>
<td>0.002</td>
</tr>
<tr>
<td>LCFA</td>
<td>0.219</td>
<td>0.001</td>
</tr>
<tr>
<td>SFA</td>
<td>0.343</td>
<td>0.002</td>
</tr>
<tr>
<td>UFA</td>
<td>0.204</td>
<td>0.001</td>
</tr>
<tr>
<td>Milk yield</td>
<td>0.262</td>
<td>0.017</td>
</tr>
<tr>
<td>Fat yield</td>
<td>0.222</td>
<td>0.002</td>
</tr>
<tr>
<td>Protein yield</td>
<td>0.216</td>
<td>0.006</td>
</tr>
<tr>
<td>Fat %</td>
<td>0.491</td>
<td>0.015</td>
</tr>
<tr>
<td>Protein %</td>
<td>0.468</td>
<td>0.007</td>
</tr>
<tr>
<td>F:P</td>
<td>0.270</td>
<td>0.001</td>
</tr>
<tr>
<td>SCS</td>
<td>0.033</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Table 4.5. Average daily genetic correlations between short-chain (SCFA), medium-chain (MCFA), long-chain (LCFA), saturated (SFA), and unsaturated (UFA) fatty acid groups and milk, fats and protein yield, fat and protein percentage, fat to protein ratio (F:P) and SCS.

<table>
<thead>
<tr>
<th></th>
<th>Milk yield</th>
<th>Fat yield</th>
<th>Protein yield</th>
<th>Fat %</th>
<th>Protein %</th>
<th>F:P</th>
<th>SCS</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCFA</td>
<td>-0.518</td>
<td>0.358</td>
<td>-0.263</td>
<td>0.863</td>
<td>0.602</td>
<td>0.750</td>
<td>0.252</td>
</tr>
<tr>
<td>MCFA</td>
<td>-0.554</td>
<td>0.451</td>
<td>-0.224</td>
<td>0.964</td>
<td>0.717</td>
<td>0.819</td>
<td>0.270</td>
</tr>
<tr>
<td>LCFA</td>
<td>-0.505</td>
<td>0.408</td>
<td>-0.192</td>
<td>0.891</td>
<td>0.680</td>
<td>0.729</td>
<td>0.246</td>
</tr>
<tr>
<td>SFA</td>
<td>-0.560</td>
<td>0.460</td>
<td>-0.237</td>
<td>0.973</td>
<td>0.707</td>
<td>0.843</td>
<td>0.274</td>
</tr>
<tr>
<td>UFA</td>
<td>-0.494</td>
<td>0.322</td>
<td>-0.228</td>
<td>0.814</td>
<td>0.617</td>
<td>0.667</td>
<td>0.237</td>
</tr>
<tr>
<td>Fat %</td>
<td>-0.620</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.734</td>
<td>-</td>
<td>0.490</td>
</tr>
</tbody>
</table>

1Genetic correlations with fatty acid groups are the average of the three analyses
Figure 4.1. Daily genetic correlations for short-chain (SCFA), medium-chain (MCFA), long-chain (LCFA), saturated (SFA), and unsaturated (UFA) fatty acid groups averaged across the three analyses, and fat percentage (FAT %) with milk yield.

Figure 4.2. Daily genetic correlations for short-chain (SCFA), medium-chain (MCFA), long-chain (LCFA), saturated (SFA), and unsaturated (UFA) fatty acid groups with fat yield.
**Figure 4.3.** Daily genetic correlations for short-chain (SCFA), medium-chain (MCFA), long-chain (LCFA), saturated (SFA), and unsaturated (UFA) fatty acid groups with fat percentage.

**Figure 4.4.** Daily genetic correlations for short-chain (SCFA), medium-chain (MCFA), long-chain (LCFA), saturated (SFA), and unsaturated (UFA) fatty acid groups with protein yield.
**Figure 4.5.** Daily genetic correlations for short-chain (SCFA), medium-chain (MCFA), long-chain (LCFA), saturated (SFA), and unsaturated (UFA) fatty acid groups and fat percentage (FAT %) with protein percentage.

**Figure 4.6.** Daily genetic correlations for short-chain (SCFA), medium-chain (MCFA), long-chain (LCFA), saturated (SFA), and unsaturated (UFA) fatty acid groups with fat to protein ratio.
Figure 4.7. Daily genetic correlations for short-chain (SCFA), medium-chain (MCFA), long-chain (LCFA), saturated (SFA), and unsaturated (UFA) fatty acid groups averaged across the three analyses and fat percentage (FAT %) with somatic cell score.
ABSTRACT
The objectives of this study were to investigate the sources of variation in milk fat globule (MFG) size in bovine milk and its prediction using mid-infrared (MIR) spectroscopy. Mean MFG size was measured in 2,076 milk samples from 399 Ayrshire, Brown Swiss, Holstein, and Jersey cows, and expressed as volume moment mean ($D_{4.3}$) and surface moment mean ($D_{3.2}$). The mid-infrared spectra of the samples and milk performance data were also recorded during routine milk recording and testing. The effects of breed, herd nested within breed, days in milk, season, milking period, age at calving, parity, and individual animal on the variation observed in MFG size were investigated. Breed, herd nested within breed, days in milk, season, and milking period significantly affected mean MFG size. Milk fat globule size was the largest at the beginning of lactation and subsequently decreased. Milk samples with the smallest MFG size on average came from Holstein cows, and the largest from Jersey and Brown Swiss cows. Partial least squares regression was used to predict MFG size from MIR spectra of samples with a calibration dataset containing 2,034 and 2,032 samples for $D_{4.3}$ and $D_{3.2}$, respectively. Coefficients of determination of cross validation for $D_{4.3}$ and $D_{3.2}$ prediction models were 0.51 and 0.54, respectively. The associated ratio of performance deviation (RPD) values were 1.43 and 1.48 for $D_{4.3}$ and $D_{3.2}$, respectively. With these models, individual mean MFG size cannot be accurately predicted, but results may be sufficient to screen samples for having either small or large MFG on average. Significant but low correlations of $D_{4.3}$ and $D_{3.2}$ with milk fat yield were estimated (0.16 and 0.21, respectively). Significant Pearson correlation coefficients for fat percent with $D_{4.3}$ and $D_{3.2}$ were observed (0.34 and 0.36, respectively). This correlation was greater between milk fat percentage and predicted MFG size than with measured MFG size with coefficients of 0.47 and 0.49 for $D_{4.3}$ and $D_{3.2}$, respectively. The MIR prediction equations are potentially overusing the correlation between fat and MFG size and exploiting the strong relationship between the MIR spectra and total milk fat. However, the predictions of MFG size are able to determine variation in mean globule size beyond what would be achieved just by looking at the correlation with fat production.
Key words: milk fat globule, mid-infrared spectroscopy

INTRODUCTION

Over 95% of the total fat in milk is present in the form of milk fat globules (MFG) (Keenan and Dylewski, 1995), a triglyceride globule surrounded by a tripartite cellular membrane. MFG are secreted in a wide variety of sizes ranging in diameter from about 0.1 to 15 µm. The average bovine MFG is approximately 2.5 to 4.6 µm in diameter although globule size shows considerable variation (Huppertz and Kelly, 2006). The size distribution and average diameter of MFG may depend on a variety of factors including breed, the individual animal, stage of lactation, diet, and season (Mulder and Walstra, 1974; Wiking et al., 2003; Carroll et al., 2006; Martini et al., 2013; Logan et al., 2014). Most studies examining variation in MFG size have used small sample size to explore these effects (< 80 cows, and cows from a single farm).

The main compositional differences in lipids between small and large MFG relate to the ratio of the MFG core to the membrane. Small MFG have more membrane material per unit of fat than large MFG. The MFG membrane is valuable to the human diet as it is naturally rich in important minor lipids and glycoproteins. Bovine MFG membrane has been suggested as a prospective nutraceutical due to the many potential health-benefits of its components (Spitsberg, 2005; Hintze et al., 2011).

Milk fat globule size also has critical implications for the technological and sensory properties of many dairy products. This trait is of particular interest in the manufacturing of cheeses, as the interaction between the surface of MFG and the casein matrix influences both the structure and texture of the finished product (Michalski et al., 2003).

A routine determination of MFG size is infeasible due to the time, cost, and the complexity of the requirement of fresh milk. Mid-infrared (MIR) technology provides an opportunity to obtain phenotypes on a considerable number of samples at a low cost, with the technology already used in regular milk recording. Recently, many researchers have been looking at using MIR spectroscopy to phenotype additional, more detailed milk composition, milk properties, and cow characteristics (De Marchi et al., 2014). It has been shown that prediction is possible for some fatty acids (Soyeurt et al., 2011), protein composition (Rutten et al., 2011), lactoferrin (Soyeurt et al., 2012), minerals (Soyeurt et al., 2009), beta-hydroxybutyrate, and acetone (De Roos et al., 2007) with varying success. Mid-infrared
spectroscopy could be used to predict mean MFG size by way of the compositional differences of samples with differing MFG sizes.

The determination of MFG size in milk samples and the factors influencing size may be of interest to the dairy industry both in terms of nutrition and manufacturing. The objectives of this study were: 1) to investigate the observed variation in mean MFG size from a large number of individual milk samples; 2) to examine the effectiveness of MIR spectroscopy in predicting mean MFG size; and 3) to determine the relationship of MFG size with routinely recorded milk production traits.

**MATERIALS AND METHODS**

*Data Collection*

A total of 44 herds with Ayrshire, Brown Swiss, Holstein, or Jersey cows that were enrolled in Canadian DHI milk recording programs were selected from across the provinces of Alberta, Ontario, and Quebec. Approximately 10 cows from each herd, 5 at the beginning of lactation and 5 in mid-lactation on the first test, were identified for milk collection. Individual milk samples (50 mL) were collected from cows multiple times throughout their lactation, and potentially their subsequent lactation, during routine milk testing by Canadian DHI partners, CanWest DHI (Guelph, ON, Canada) for Alberta and Ontario herds and Valacta (Ste-Anne-de-Bellevue, QC, Canada) for Quebec herds. Collection occurred between March 2013 and October 2014 for CanWest DHI and between December 2013 and May 2015 for Valacta.

Milk samples were transported to either a CanWest DHI or Valacta milk laboratory. A portion of the milk sample was removed and analyzed by a MIR spectrometer (MilkoScan FT6000; FOSS, Hillerød, Denmark) using standard milk recording procedures. Spectra were collected from two machines, one at CanWest DHI and one at Valacta. The MIR data for each sample contained 1,060 data points in the infrared range of 5,000 to 900 cm\(^{-1}\). At the same laboratories, fat and protein content were determined from the spectra. Standardization of the historical spectra between the two machines and across time was performed as previously described by Bonfatti et al. (2016).

The remaining quantity of the milk sample was sent fresh (never frozen) from the DHI laboratory to the University of Guelph (Guelph, ON, Canada) for MFG measurement. Milk samples were analyzed between 1 and 21 days after collection (on average within 5.5 days from
the collection). MFG size distribution was measured by integrated light scattering with a Malvern Mastersizer 2000 (Southborough, MA, USA). Immediately before measurement, samples were diluted 1:1 in 80 mM of EDTA/NaOH solution (pH 7) to minimize the signal from the casein micelles by adding 1 mL of EDTA buffer to 1 mL of milk and vortexing. The absorption coefficients of liquid milk fat used for measurement were $0.5 \times 10^{-5}$ and $1.7 \times 10^{-5}$ at 633 and 466 nm, respectively. The refractive indexes of MFG in water were 1.458 and 1.460 at 633 and 466 nm, respectively. Whole milk was measured at 2,800 rpm speed stirring. The obscuration was 10% to 15%. The diameter of the MFG was recorded as volume moment mean ($D_{[4,3]}$), and surface moment mean ($D_{[3,2]}$) defined by the equation:

$$d[k, z] = \frac{\sum N_i d_i^k}{\sum N_i d_i^z},$$

where $N_i$ is the number of globules in a size class of $d_i$.

At the conclusion of collection, there were 2,083 milk samples with mean MFG size measured from 392 cows representing the four breeds with saved MIR spectra. The number of samples per cow ranged from 1 to 14 and averaged 5.31 samples per cow. There were 26 cows with only one MFG size record. Sampling numbers varied greatly due to the number of times a herd was visited, sample spoiling or freezing during transport, cow health complications, cows entering their dry period, and cow removal from the herd. Two samples were removed because their $D_{[3,2]}$ measurement was greater than that for $D_{[4,3]}$ and a recording error was assumed. An additional five records with values greater than 4 standard deviations from the mean were removed, in order to eliminate measurement errors and milk samples where MFG aggregation likely occurred. This left 2,076 samples in the dataset. The number of cows, samples, and herds with measured MFG size and their descriptive statistics are shown in Table 5.1. The selection of herds was geared toward obtaining as much variation in the fine milk components as possible, which included finding herds employing a wide range of management practices. Many of these herds had Holsteins, the most prominent dairy breed in Canada, and thus the majority of the samples have come from Holstein cows.

**Phenotypic Analysis**

Complete information on the test day and the cow was available for 1,826 of the analyzed samples coming from 361 animals in 42 herds. Information included herd, breed, days in milk,
season, milking period (AM or PM), age at calving, and parity. Data were analyzed using the MIXED procedure of SAS (SAS Institute, 2013), with a repeated statement to allow for multiple records per cow using the following linear model:

\[ y_{ijklmno} = \mu + H_i(B)_j + B_j + DIM_k + S_l + MP_m + a_n + e_{ijklmno}, \]

where \( y_{ijklmno} \) is the dependent variable D[4,3] or D[3,2]; \( \mu \) is the overall mean; \( H_i(B)_j \) is the fixed herd effect nested within breed (\( i = 1 \) to 42); \( B_j \) is the fixed effect of breed (\( j = 1 \) to 4), \( DIM_k \) is the fixed effect of days in milk class (\( k = 1 \) to 11; class 1: 5 - 30 d, class 2: 31 - 60 d, class 3: 61 – 90 d, … , class 10: 271 – 305 d, and class 11: > 305 d after calving), \( S_l \) is the fixed effect of season (\( l = 1 \) to 4; class 1: January-March, class 2: April-June, class 3: July-September, class 4: October-December); \( MP_m \) is the fixed effect of milking period (\( m = 1 \) to 2; class 1: AM, class 2: PM), \( a_n \) is the random effect of animal (\( n = 1 \) to 361); \( e_{ijklmno} \) is the random error.

No relationship matrix was used and animals were therefore assumed to be unrelated. Parity and age at calving did not have a significant effect on MFG size (\( P = 0.87 \) and \( P = 0.60 \), respectively) and were excluded from the final model. A significance level of \( P < 0.05 \) was used. Least squares means (LSM) were compared by use of Scheffé’s adjustment to determine significant differences between the factor levels. Orthogonal contrasts were estimated between LSM for the effects of DIM (linear, quadratic, cubic, and quartic component).

**Predictive MIR Model**

Prediction equations were obtained by partial least squares regression using PLS procedures of SAS (SAS Institute, 2013). Regions of the MIR spectra from 3,105 to 3,444 cm\(^{-1}\) and 1,628 to 1,658 cm\(^{-1}\), characterized by low signal to noise ratio caused by the high absorption of water, were not used for the PLS analysis. Thus 862 spectral variables were used. No further spectral pre-treatments were used to develop the prediction equations. Milk sample spectral outliers were assessed and removed during the PLS procedure. Originally, all samples with both MFG size and spectral data available (\( N = 2,076 \)) were included in the calibration set and models were created. The root mean square error of the standardized predictors was examined for each sample as a measure of the distance between the data point and the model plane in X-space. Sample spectra with a mean square error greater than 3 standard deviations above the mean value were considered outliers and omitted from the analysis. The PLS procedures were then used on the remaining data (\( N = 2,034 \) and 2,032 for D[4,3] and D[3,2], respectively) to produce the final
calibration equation. Full cross-validation was utilized with a leave-one-out cross-validation procedure, whereby one sample is removed and the remaining samples are used to produce the model and predict the discarded sample. The process is repeated until each sample has been predicted once, with the validation errors saved each time to produce the standard error of cross-validation (SECV). In order to assess the accuracy and utility of the calibration equations, the coefficient of determination of cross-validation ($R^2_{cv}$), the ratio of performance deviation (RPD), and the range error ratio (RER) were calculated. For RPD, the ratio of SD to SECV, a higher value is desired, with a RPD greater than 2 enabling good predictions (De Marchi et al., 2011). The RER is calculated by dividing the range of the reference data by the SECV, and can be used to determine the practical application of the prediction equations. Calibration equations with a RER value between 7 and 20 are considered poor to fair and may be adequate for screening purposes, and values between 21 and 40 are good to very good and more suitable for quality and process control applications (Williams, 2001).

**Correlation with Production Traits**

Milk production from milk recording and fat and protein yields calculated routinely from MIR spectra by DHI labs were available for 1,957 of the samples used in the MFG calibration dataset. Production traits were measured on the same test-day milk sample as the MFG size analysis was performed. Correlation coefficients were estimated with SAS among the measured and predicted measures of MFG size along with milk yield, fat yield and percentage, and protein yield and percentage. Correlations with $P$-values less than 0.05 were deemed significant.

**RESULTS AND DISCUSSION**

**Phenotypic Analysis**

Descriptive statistics of measured MFG size are summarized in Table 5.1. The average $D[4,3]$ for all samples analyzed was 4.24 µm with a standard deviation of 0.55 µm and the average $D[3,2]$ was 3.53 µm with a standard deviation of 0.35 µm. The herd nested within breed, breed, days in milk, season, and milking period all had a significant effect on MFG size ($P < 0.05$). An effect of milking period was found with evening milk having significantly larger MFG on average ($P < 0.05$). Least squares means for $D[4,3]$ were 4.65 µm for morning and 4.68 µm for evening milk and, for $D[3,2]$, the LSM were 3.74 µm for morning and 3.78 µm for evening
milk. In this study, information was only available on the time of day of milking and not milking frequency or interval between milkings. The categorization of this variable may be misguided if participating herds use unequal intervals between milkings, which is a factor likely to have an effect as observed by Wiking et al. (2006).

There was a significant effect ($P < 0.001$) of the seasons on MFG size. For D[4,3] and D[3,2], the milk collected in the spring (April to June) had larger MFG on average than milk collected in the later half of the year (Figure 5.1). Also, for D[4,3] only, MFG size was larger in the first three months compared to the last three months of the year. The large MFG in spring milk in this study is contrasting to Mulder and Walstra et al. (1974) who noted smaller MFG in spring milk compared to winter milk. Couvreur et al. (2006) observed a decrease in mean MFG size when a greater amount of fresh grass was included in the cow’s diet, which could increase in the spring and summer. Diet and management practices of the contributing herds were not known, and operations may have similar feeding through the entire year. In addition, the spring season experienced in the present study may not be comparable, as primarily pasture feeding may not commence until much later in the year in some of the participating regions of the country.

Milk from Jersey and Brown Swiss cows had on average larger MFG than Ayrshire and Holstein cows, with Holstein having on average the smallest (Figure 5.2). Carroll et al. (2006) did find MFG from Jersey cows were larger in diameter on average than that for Brown Swiss cows, as well as Holstein cows. Mulder and Walstra (1974) also noted that milk from Jersey, as well as Guernsey cows, contained larger MFG on average than milk from Holstein-Friesian type cows.

Milk fat globule size was greatest at the start of lactation and generally decreased thereafter (Figure 5.3). In the present study, the relationship between both D[4,3] and D[3,2] with DIM was determined to be cubic in nature ($P < 0.001$). Maximum MFG size at the beginning of lactation, followed by a decrease was also noted by Mulder and Walstra (1974) and Wiking et al. (2003). The association between the effects of breed and days in milk with MFG size may be connected to relationship of breed and days in milk with fat yield.
MFG Calibration Equations

Descriptive statistics of the samples used to create the calibration equations, the predicted values, and the studied milk-quality traits are summarized in Table 5.2. The fitting statistics of the predictive models for D[4,3] and D[3,2] are shown in Table 5.3. The model for D[3,2] predicted slightly better with an $R^2_{cv}$ of 0.54 compared to 0.51 for D[4,3]. The corresponding RPD values of D[4,3] and D[3,2] prediction models were 1.43 and 1.48, respectively, showing that the models are doing a relatively poor job at accurately predicting mean MFG size. The RER values, which are both greater than the 7 stated by Williams (2001) as a minimum for screening purposes, also indicate that, although the models are poor, they do explain enough of the variation in MFG size to have some gauge of the average MFG size. Figure 5.4 displays the measured vs. predicted values of D[4,3] and D[3,2]. The current MIR prediction equations do not quantify the size adequately but are sufficient to screen milk samples and identify milk with large or small MFG on average. A relationship between the measured and predicted values can be noted and generally milk samples with smaller MFG on average predict in the lower part of the range and samples with the larger MFG on average predict in the high part of the range. The standard deviation calculated for the predicted values of D[4,3] and D[3,2] were less than found in the measured values (Table 5.2). The models tend to predict MFG size toward the means, with small MFG samples predicting higher and large MFG predicted lower than their true values. The slightly better predictive ability of the D[3,2] measure makes it more appealing since the two traits are very similar in nature. In addition, D[3,2] is more relevant where specific surface area is meaningful such as with the MFG membrane in improving the nutritional value of milk and milk products.

The capability of predicting milk components and properties from their MIR spectra has had variable results depending on the trait. Much success has been achieved in using MIR spectra to predict individual and groups of fatty acids. Soyeurt et al. (2011) reported $R^2$ values greater than 0.95 for saturated, unsaturated, short, medium and long-chain fatty acid groups. The predictive ability of individual fatty acids is generally better for major fatty acids with a relationship between $R^2$ and concentration observed (De Marchi et al., 2011). Differences in fine fat composition are thus detectable in milk samples using MIR spectroscopy and MFG size prediction equations may draw on this information.
The intent in predicting MFG size from MIR spectra was to exploit the compositional differences exhibited between milk with small and large MFG and not directly detect particle size. In the process of procuring the MIR spectra of a milk sample, a homogenizer is added to disrupt the MFG and make them smaller. This is necessary because MIR radiation has limited penetration depth, and therefore, the spectra are sensitive to the presence of MFG or fat biofilms and scattering would occur. The major compositional difference between differently sized MFG is the relative concentration of phospholipid, as well as other membrane materials, to triacylglycerol. The ratio of surface area to volume of a sphere is $3 \div \text{radius}$, and consequently, compositional changes related to the ratio of membrane to core material with increasing size is a more complex relationship to predict from MIR spectra.

Other compositional differences have been noted between small and large MFG, primarily fatty acid concentration, although many disparities exist likely due to large differences in methodology. Average MFG diameter was found to be positively correlated with C16:0, C16:1, C18:0 and C18:1 but not with C4:0 to C14:0, C18:2, or C18:3 by Wiking et al. (2004). Conversely, Briard et al. (2003) discovered higher amounts of C18:1 and C18:2 in small MFG than in large MFG. Lopez et al. (2011) did not find differences in the amount of total saturated and unsaturated fatty acids or in the amounts of short chain fatty acids (C6:0 to C10:0) between different sized MFG. They did, however, find significantly more C12:0, C14:0 C16:0, C18:1 trans, and C18:2c9tr11 and less C18:0 and C18:1c9 in small MFG than large MFG. Conversely, Timmen and Patton (1988) found less short-chain fatty acids (C4:0 to C10:0) in small MFG compared to large MFG, but did also find less C18:0 and greater C18:1. It is, therefore, important to note that the nature and magnitude of compositional changes observed between small and large MFG may not be consistent and can be influenced by the diet (Wiking et al., 2004; Lopez et al., 2008), breed (Gallier et al., 2011), and season (Briard et al., 2003). This may limit the suitability of using MIR technology to predict MFG size through compositional differences and a reason for the low coefficient of determination. Furthermore, in this study mean diameter was the only measure available and no other valuation of the distribution of MFG size for a milk sample was known. The overall fat composition of a sample is liable to change depending on the shape of the MFG size distribution.
**Correlation with Production Traits**

Pearson correlation coefficients amongst measured and predicted MFG size, and production traits are shown in Table 5.4. A high correlation of 0.90 was found between the two measures of MFG size. This relationship was expected as the two parameters are describing much the same trait, but the assessment of D[4,3] is more sensitive to the presence of large particles and D[3,2] is more sensitive to small particles in the size distribution. A low but significant correlation was found between MFG size and protein percent ($P < 0.001$ and $P = 0.04$ for D[4,3] and D[3,2], respectively), but no significant correlation existed with protein yield ($P = 0.14$ and $P = 0.72$, respectively). The relationship between milk yield, protein yield, and protein percent with MFG size was largely unchanged between the measured and predicted MFG size measures. A significant correlation ($P = 0.04$) was observed between milk yield and predicted D[4,3] that was not present between milk yield and D[4,3]. However, this correlation was still very weak (-0.05).

A significant correlation ($P < 0.001$) between fat yield and measured D[4,3] and D[3,2] size of 0.16 and 0.21, respectively, was found. The correlation between fat percentage and D[4,3] and D[3,2] was greater at 0.34 and 0.36, respectively. Wiking et al. (2004) also found D[4,3] was explained partially by diurnal fat production in Danish Holstein cows ($R^2 = 0.54$). They theorized that MFG membrane material synthesis might not be able to increase as fat yield increases and as a result fat droplets grow larger before they are enveloped with plasma membrane in the secretory apical membrane. Larger MFG could be secreted to reduce the amount of membrane lost per unit volume of fat when membrane material is limited. However, there is currently little understanding of how the eventual size of the MFG is decided during their formation and secretion, and whether fat production has a causative association with MFG size. The positive connection between MFG size and fat production has also been investigated through animal diet. When cows were fed concentrate with a high quantity of saturated lipids, Wiking et al. (2003) found the milk had a greater fat content and significantly larger MFG than milk from cows fed other diets. This was also demonstrated by Carroll et al. (2006) who observed with 12 cows from 3 dairy breeds that average MFG diameter tended to increase with increased fat production as a result of feeding diets with increased fat.

The correlation coefficients of the MFG size predictions with fat yield and fat percent were greater than that of their measured equivalents. The correlation coefficients for fat yield
increased to 0.23 and 0.30 for predicted D[4,3] and D[3,2], respectively. Higher correlations of 0.47 for predicted D[4,3] and 0.49 for predicted D[3,2] with fat percent were found. The increase in the correlation coefficient suggested that the predictive model, for both measures of MFG size, are likely drawing on the correlation between MFG size and fat content, and using absorption bands more associated to total fat. However, the correlation coefficient between the measured and predicted D[4,3] and D[3,2] were 0.72 and 0.74, respectively, which are greater than the correlation with fat and, thus, are better indicators of MFG size.

Prediction equations are only beneficial if the model is in fact using information directly related to the trait or component. The PLS methods used to create the predictions for milk traits from MIR spectroscopy depend on the correlations between the MIR absorption and the trait, but models may also inadvertently be built on indirect correlations between the trait of interest and absorption bands not directly caused by that trait. If these indirect correlations are not preserved, the predictions will lose accuracy and become ineffective. Eskildsen et al. (2014) suggested that MIR predictions of individual fatty acids might not be valuable since the models are providing information related to total fat rather than the individual fatty acid and future samples may predict poorly if these correlations are inconsistent. The compositional differences between milk with varying MFG size is complex and may range as discussed above. Accordingly, it is conceivable that the predictive models may be using information indirectly correlated to components such as total fat content.

The measured and predicted values of MFG size at a common fat content (4%; the mode of the data) were investigated more closely, and predicted MFG size still showed variation similar to the measured values (Figure 5.5). For a given fat content, milk samples with small MFG on average had a lower predicted size than samples with measured large MFG. Although the MIR prediction exploits the correlation with milk fat, the prediction does apply information beyond milk as small and large MFG size on average can still be differentiated to some extent for a given fat content.

CONCLUSIONS

The present study investigated the variation observed in mean MFG size and MIR prediction as a potential tool for the screening of MFG size. Mean MFG size does vary between breeds, days in milk, season, milk period, and individual animal. The current MIR prediction
equations are poor for precise quantification, but may be useful to differentiate samples with small and large MFG on average. The increased strong correlation between predicted MFG size and milk fat content demonstrates the models are using correlations not truly indicative of MFG size. However, predicted MFG size still shows variation within milk fat categories. The further analysis of the milk samples for additional minor milk components will give further insight into potential compositional changes in milk with differing MFG size that MIR models may be employing in their predictions.
Table 5.1. Number of cows, records, and herds and the descriptive statistics of measured milk fat globule volume moment mean (D[4,3]) and surface moment mean (D[3,2]) for each breed.

<table>
<thead>
<tr>
<th>Breed</th>
<th>Cows</th>
<th>Records</th>
<th>Herds(^1)</th>
<th>Mean (µm)</th>
<th>SD (µm)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>D[4,3]</td>
<td>D[3,2]</td>
</tr>
<tr>
<td>Ayrshire</td>
<td>55</td>
<td>194</td>
<td>7</td>
<td>4.37</td>
<td>3.69</td>
</tr>
<tr>
<td>Brown Swiss</td>
<td>23</td>
<td>97</td>
<td>3</td>
<td>4.23</td>
<td>3.61</td>
</tr>
<tr>
<td>Holstein</td>
<td>254</td>
<td>1,584</td>
<td>29</td>
<td>4.14</td>
<td>3.47</td>
</tr>
<tr>
<td>Jersey</td>
<td>60</td>
<td>201</td>
<td>8</td>
<td>4.90</td>
<td>3.88</td>
</tr>
<tr>
<td>Total</td>
<td>392</td>
<td>2,076</td>
<td>44</td>
<td>4.24</td>
<td>3.54</td>
</tr>
</tbody>
</table>

\(^1\)Three herds had multiple breeds

Table 5.2. Descriptive statistics of measured milk fat globule volume moment mean (D[4,3]) and surface moment mean (D[3,2]) used in prediction models, predicted milk fat globules volume moment mean (pD[4,3]) and surface moment mean (pD[3,2]), and milk production traits.

<table>
<thead>
<tr>
<th>Trait</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>D[4,3] (µm)</td>
<td>2,034</td>
<td>4.24</td>
<td>0.55</td>
<td>2.78</td>
<td>6.35</td>
</tr>
<tr>
<td>D[3,2] (µm)</td>
<td>2,032</td>
<td>3.53</td>
<td>0.35</td>
<td>2.21</td>
<td>4.73</td>
</tr>
<tr>
<td>pD[4,3] (µm)</td>
<td>2,034</td>
<td>4.24</td>
<td>0.40</td>
<td>3.17</td>
<td>6.34</td>
</tr>
<tr>
<td>pD[3,2] (µm)</td>
<td>2,032</td>
<td>3.53</td>
<td>0.26</td>
<td>2.75</td>
<td>4.67</td>
</tr>
<tr>
<td>Fat (kg)</td>
<td>1,957</td>
<td>1.21</td>
<td>0.38</td>
<td>0.20</td>
<td>3.53</td>
</tr>
<tr>
<td>Fat (%)</td>
<td>1,957</td>
<td>3.98</td>
<td>0.75</td>
<td>1.34</td>
<td>8.03</td>
</tr>
<tr>
<td>Milk (kg)</td>
<td>1,957</td>
<td>31.46</td>
<td>10.78</td>
<td>5.00</td>
<td>72.60</td>
</tr>
<tr>
<td>Protein (kg)</td>
<td>1,957</td>
<td>1.03</td>
<td>0.29</td>
<td>0.20</td>
<td>2.06</td>
</tr>
<tr>
<td>Protein (%)</td>
<td>1,957</td>
<td>3.35</td>
<td>0.40</td>
<td>2.42</td>
<td>4.97</td>
</tr>
</tbody>
</table>
Table 5.3. Fitting statistics of prediction models for milk fat globule volume moment mean (D[4,3]) and surface moment mean (D[3,2])

<table>
<thead>
<tr>
<th>Trait</th>
<th>N</th>
<th>SECV</th>
<th>R²cv</th>
<th>RPD</th>
<th>RER</th>
</tr>
</thead>
<tbody>
<tr>
<td>D[4,3]</td>
<td>2,034</td>
<td>0.39</td>
<td>0.51</td>
<td>1.43</td>
<td>9.26</td>
</tr>
<tr>
<td>D[3,2]</td>
<td>2,032</td>
<td>0.24</td>
<td>0.54</td>
<td>1.48</td>
<td>10.55</td>
</tr>
</tbody>
</table>

¹Standard Error of Cross Validation (SECV), Coefficient of determination of cross validation (R²cv), the ratio of performance deviation (RPD), and the range error ratio (RER)

Table 5.4. Pearson correlation coefficients between measured milk fat globule volume moment mean (D[4,3]) and surface moment mean (D[3,2]); predicted milk fat globule volume moment mean (pD[4,3]) and surface moment mean (pD[3,2]); and milk production traits.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>D[4,3]</td>
<td>0.90**</td>
<td>0.72***</td>
<td>0.68***</td>
<td>0.16**</td>
<td>0.34***</td>
<td>-0.03</td>
<td>-0.03</td>
<td>0.08***</td>
</tr>
<tr>
<td>D[3,2]</td>
<td>0.70***</td>
<td>0.74***</td>
<td>0.21***</td>
<td>0.36***</td>
<td>-0.01</td>
<td>-0.01</td>
<td>0.01</td>
<td>0.05*</td>
</tr>
<tr>
<td>pD[4,3]</td>
<td>0.96***</td>
<td>0.74***</td>
<td>0.23***</td>
<td>0.47***</td>
<td>-0.05*</td>
<td>-0.03</td>
<td>0.10***</td>
<td></td>
</tr>
<tr>
<td>pD[3,2]</td>
<td></td>
<td>0.30***</td>
<td>0.49***</td>
<td>-0.01</td>
<td>0.01</td>
<td>0.05*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*P < 0.05; **P < 0.001.
Figure 5.1. Least squares means of milk fat globule size measured as (a) volume moment mean (D[4,3]), and (b) surface moment mean (D[3,2]) for season. Means with different superscript letters indicate $P < 0.05$. 
Figure 5.2. Least squares means of milk fat globule size measured as (a) volume moment mean (D[4,3]), and (b) surface moment mean (D[3,2]) for Holstein (HO; N = 1,406 samples), Ayrshire (AY; N = 171), Brown Swiss (BS; N = 60), and Jersey (JE; N = 189) breeds. Means with different superscript letters indicate $P < 0.05$. 
Figure 5.3. Least squares means of milk fat globule size measured as volume moment mean (D[4,3]), and surface moment mean (D[3,2]) for classes of days in milk.
Figure 5.4. Predicted vs. measured milk fat globule size measured as (a) volume moment mean (D[4,3]), and (b) surface moment mean (D[3,2]).
Figure 5.5. Predicted vs. measured milk fat globule size measured as surface moment mean (D[3,2]) for samples with a common (4.0%) milk fat content.
Chapter 6

GENETIC PARAMETERS OF BOVINE MILK FAT GLOBULE SIZE, MID-INFRARED PREDICTED FAT GLOBULE SIZE, AND FAT AND PROTEIN CONTENTS

ABSTRACT

The objective of this study was to estimate the heritability of milk fat globule (MFG) size and mid-infrared predicted MFG size, and the genetic correlations among them and with milk fat and protein percentage in Holstein cattle. Average MFG size was measured in 1,583 milk samples taken from 254 Holstein cows from 29 herds across Canada. Size was expressed as volume moment mean (D[4,3]) and surface moment mean (D[3,2]). Analyzed milk samples also had MIR-predicted average MFG size, and fat and protein percentages were obtained for all test-day milk samples in the cow’s lactation. Univariate and bivariate repeatability animal models were used to estimate heritability and genetic correlations. Moderate heritabilities of 0.392 (0.183) and 0.494 (0.185) were found for D[4,3] and D[3,2], respectively, and a strong genetic correlation was found between the two traits (0.975). The heritabilities for the MIR-predicted MFG size were lower than those estimated for the measured MFG size at 0.344 for predicted D[4,3] and 0.301 for predicted D[3,2]. The genetic correlation between measured and predicted D[4,3] was 0.721, and slightly higher between measured and predicted D[3,2] at 0.792, likely due to the better prediction accuracy of D[3,2]. Milk fat percentage had moderate genetic correlations with both D[4,3] and D[3,2] (0.537 and 0.692, respectively). The predicted MFG size had much stronger genetic correlations with fat percentage (greater than 0.97 for both predicted D[4,3] and D[3,2]). This suggests a limitation for the use of the predicted values of MFG size as indicator traits for true average MFG size in milk. Larger samples sizes are required to provide better evidence of the estimated genetic parameters but there appears to be a genetic component to the average MFG size in bovine milk and the variation could be exploited in selection programs.

Keywords: Milk fat globule, mid-infrared prediction, genetic parameter

INTRODUCTION

Fat is an important component of milk as it has a role in the economics, nutrition, and technological properties of milk and milk products. More than 95% of the total fat in milk is present in the form of milk fat globules (MFG) (Keenan and Dylewski, 1995), a triple layer
cellular membrane surrounding a globule of primarily triacylglycerol. Walstra (1969) reported MFG with a diameter less than 1 µm represented 80% of all MFG, but only 5% of milk fat volume, while MFG ranging from 1 to 8 µm comprised approximately 94% of fat. The remaining 1 to 2% of the fat volume comes from a small number of large MFG with a diameter greater than 8 µm. Milk fat globule triacylglycerols are synthesized at the rough endoplasmic reticulum, accumulated, and then released as microlipid droplets into the cytosol with a surface coat of polar lipids and proteins derived from the endoplasmic reticulum (Mather and Keenan, 1998). The droplets grow in size through coalescence with each other as they migrate from their sites of origin, to the apical region, where growth is most pronounced (Stemberger and Patton, 1981). The mechanisms involved in the fusion of lipid droplets and whether these processes are regulated or random remain speculative. The average size of MFG can be affected by multiple factors including breed (Carroll et al., 2006), stage of lactation (Mulder and Walstra, 1974), and diet (Wiking et al., 2003). A potential genetic regulation of MFG size has been suggested as well (Campbell, 1932). A variation in MFG diameter of up to 1 µm between individual animals was reported by Mulder and Walstra (1974), suggesting the possibility of selecting cows for the size of their MFG.

Milk fat globule size is a trait of interest because it affects the technological and sensory properties of milk and milk products. Moisture content and texture of cheese can also be altered by using milk with either small or large MFG (Michalski et al., 2003). Of interest to consumers, are the health benefits of many MFG membrane components, which includes important minor lipids and glycoproteins (Spitsberg, 2005). At a given amount of fat, cows producing small MFG will be producing higher amounts of the beneficial MFG membrane material.

Mid-infrared (MIR) spectroscopy has been used to quantify fat and protein content in milk samples in dairy herd improvement for decades. The use of this technology has more recently expanded to predict additional milk components and properties and cow physiology, including fatty acid content (Rutten et al., 2009; Soyeurt et al., 2011), protein composition (Bonfatti et al., 2011), mineral content (Soyeurt et al., 2009), milk coagulation properties (Cecchinato et al., 2009), and energy intake and efficiency (McParland et al., 2014). The MIR predicted values could act as indicator traits to allow for the routine genetic evaluation of many meaningful traits for which large-scale phenotyping would otherwise be unattainable. The adequacy of MIR predicted traits for selection purposes depends on the performance of the
prediction equation. Soyeurt et al. (2011) proposed that prediction models with $R^2_{cv}$ greater than 0.75 could be used for selection; however, Cecchinato et al. (2009) investigated the genetic correlations and selection response and deduced that predictions below this threshold may still be effective.

Few studies have investigated the genetic influence on MFG size and the corresponding estimated genetic parameters. The objective of this study was to estimate genetic parameters of measured and MIR-predicted average MFG size and their genetic correlations with each other along with milk fat and protein percentages.

**MATERIALS AND METHODS**

*Data*

Milk samples were collected from 254 Holstein cows from 29 herds over a 26-month period during routine milk recording by Canadian DHI partners, CanWest DHI (Guelph, ON, Canada) and Valacta (Ste-Anne-de-Bellevue, QC, Canada). At the initiation of milk sample collection, approximately ten cows per participating herd were selected, half at the beginning of their lactation and half at mid-lactation. Sampling occurred multiple times during the project timeframe for the chosen cows spanning one ($N = 131$ cows) or two lactations ($N = 123$ cows) depending on the stage of lactation at the commencement of the sampling and their ability to remain in the herd. In total, 1,589 milk samples were analyzed for average MFG size. The number of milk samples per cow ranged between 1 and 14 with 246 cows having more than one sample. On average there were 6.3 milk samples taken per cow across the collection time period. MFG size distribution was measured on fresh (never frozen) milk samples by integrated light scattering with a Malvern Mastersizer 2000 (Southborough, MA, USA) using a two-wavelength protocol. Immediately before measurement, samples were diluted 1:1 in 80 mM of EDTA/NaOH solution (pH 7) to minimize the signal from the casein micelles by adding 1 mL of EDTA buffer to 1 mL of milk and vortexing. The absorption coefficients of liquid milk fat used for measurement were $0.5 \times 10^{-5}$ and $1.7 \times 10^{-5}$ at 633 nm and 466 nm, respectively. The refractive indices of MFG in water were 1.458 and 1.460 at 633 nm and 466 nm, respectively. Whole milk was measured at 2,800-rpm speed stirring. The obscuration was 10% to 15%. The mean diameter of the MFG was calculated by the integrated software and expressed as volume moment mean ($D[4,3]$), and surface moment mean ($D[3,2]$) defined by the equation:
where $N_i$ is the number of globules in a size class of $d_i$.

If $D[4,3]$ was lower than $D[3,2]$, a recording error was assumed and the corresponding records were omitted. Records for which $D[4,3]$ or $D[3,2]$ were farther than four standard deviations away from the mean were also removed. After edits there were 1,583 milk samples with $D[4,3]$ and 1,586 with $D[3,2]$ data.

Predicted $D[4,3]$ ($pD[4,3]$) and predicted $D[3,2]$ ($pD[3,2]$) were obtained for these data from the saved MIR spectra using the calibration equations previously described by Fleming et al. (2016b). MIR calibration equations were developed to predict average MFG size in milk samples using a multi-breeding calibration set that included the Holstein samples utilized in the present study. The prediction equations performed relatively poorly with models for $D[4,3]$ and $D[3,2]$ having coefficients of determination of cross validation ($R^2_{cv}$) of 0.51 and 0.54, respectively, but may still have some utility (Fleming et al., 2016b). As a result of some milk samples not having recorded MIR spectra or having spectra deemed to be an outlier, there were 1,551 samples with $pD[3,2]$ and 1,549 with $pD[3,2]$.

Test-day fat and protein percentages determined during milk recording procedures were obtained from the Canadian Dairy Network (Guelph, ON, Canada). If one test-day in a cow’s lactation had MFG size recorded, the fat and protein percentage for all test days in that lactation were kept. As a result, there were 3,643 fat and protein percentage records kept for these traits (average of 14.5 samples per cow). A full pedigree going back as many generations as available containing 22,819 individuals was obtained. The 254 cows with records were the progeny of 169 different sires.

**Genetic Analysis**

Univariate and bivariate repeatability animal models were used to analyze the data. The genetic correlation between the repeated measurements was assumed to be one across the lactation and parities. Variance components were estimated using the average information-restricted maximum likelihood (AI-REML) procedure in the DMU package (Madsen and Jensen, 2008). The following linear animal model was applied to all traits:

$$y_{ijklmn} = \mu + HD_i + P_j + DIM_k + p_{ij} + a_m + e_{ijklmn},$$
where \( y_{ijklmn} \) is the dependent \( D[4,3] \), \( D[3,2] \), \( pD[4,3] \), \( pD[3,2] \), fat percentage, or protein percentage, \( \mu \) is the overall mean; \( HD_i \) is the fixed herd test day effect (\( i = 1 \) to 766); \( P_j \) is the fixed effect of parity (\( j = 1 \) to 4 and more), \( DIM_k \) is the fixed effect of days in milk class (\( k = 1 \) to 11; class 1: 5 – 30 d, class 2: 31 – 60 d, class 3: 61 – 90 d, ..., class 10: 271 – 305 d, and class 11: > 305 d after calving), \( pe_l \) is the random permanent environment effect (\( l = 1 \) to 254); \( a_m \) is the random additive genetic effect of the animal (\( m = 1 \) to 22,819), and \( e_{ijklmn} \) is the random error term.

Heritability was calculated using variance components estimated from the univariate analysis of each trait as, \( \sigma_a^2/(\sigma_a^2 + \sigma_{pe}^2 + \sigma_e^2) \), where \( \sigma_a^2 \) is the additive genetic variance, \( \sigma_{pe}^2 \) is the permanent environmental variance, and \( \sigma_e^2 \) is the residual variance. Repeatability was defined as \( (\sigma_a^2 + \sigma_{pe}^2)/(\sigma_a^2 + \sigma_{pe}^2 + \sigma_e^2) \). Genetic correlations between traits were estimated from the series of fourteen bivariate analyses with \( D[4,3] \), \( D[3,2] \), \( pD[4,3] \), and \( pD[3,2] \) amongst themselves and with fat and protein percentages.

**RESULTS AND DISCUSSION**

**Descriptive Statistics**

Descriptive statistics for all examined traits in the edited dataset are displayed in Table 6.1. The mean \( D[4,3] \) was 4.130 ± 0.483 µm while \( D[3,2] \) had a lower mean of 3.468 ± 0.339 µm. These values are within the range reported by (Huppertz and Kelly, 2006). The MIR predicted values for these traits had similar means to the measured MFG but the variation observed in the predicted traits was lower. This phenomenon is a consequence of the low predictive power of the model in tails whereby values in the extreme parts of the distribution (i.e. in the tails) are predicted to be closer to the mean than their corresponding true measured values.

**Heritabilities and Repeatabilities**

Heritability estimates calculated from the univariate analysis are shown in Table 6.2. The fat and protein percentages had heritabilities of 0.426 and 0.299, respectively. Miglior et al. (2007) reported for Canadian Holstein cattle heritabilities ranging from 0.533 to 0.555 for fat percentage and 0.561 to 0.586 for protein percentage. These heritabilities are greater than those found in the current study, but when considering the standard errors only the heritability of protein percent was lower than expected. The observed differences are attributable to the small
sample size and simple genetic models used in the present study. Moderate heritabilities of 0.392 and 0.494 were found for D[4,3] and D[3,2], respectively. This suggests that the genetics of the cow plays a significant role in determining the average size of the MFG that are formed and secreted in milk, and selection for this trait is possible. There are limited studies examining the heritability of average MFG size in milk. Hassaneyn (1965) reported a high heritability of 0.73 for butterfat globule size using a half-sib method in German Friesian cows, although out of date and dissimilar methodologies were used. Regardless, this finding still supports the notion that MFG size is heritable in dairy cattle. The variation in MFG size observed by Couvreur et al. (2007) and Logan et al. (2014) for cows within a herd may thus be explained in part by genetic differences with or without other unaccounted for on-farm factors. Argov-Argaman et al. (2013) used the ratio of phospholipid to triacylglycerol as an indicator of MFG size and discovered a significant relationship between this ratio and diacylglycerol acyltransferase 1 (DGAT1) genotype. DGAT1 is involved in the last step of the synthesis of triacylglycerol and has a major effect on milk fat content (Grisart et al., 2002). This gene could increase the average size of MFG if the gene variant resulting in increased triacylglycerols, the main component of the MFG core, is present. Although the results of the present study provide evidence that the average MFG size in a milk sample has a genetic component, the complete genetic architecture of MFG size requires further investigation. The regulation of the overall size distribution of MFG within a milk sample and how the size of any one particular MFG is determined remains unclear. The process by which the mature size of an individual MFG is determined as lipid droplets are transported, coalesced, and secreted is perplexing and requires a comprehensive examination to see the role, if any, of specific genes.

Repeatability estimates were similar for all examined traits (Table 6.1). Repeatabilities of 0.579 and 0.623 were estimated for D[4,3] and D[3,2], respectively. Slightly lower and closer to each other repeatabilities were found for pD[4,3] (0.550) and pD[3,2] (0.559). Genetic evaluation of MFG size would therefore benefit from multiple measurements throughout the lactating lifetime of a cow. Since predicted MFG size is coming from the MIR spectra captured during routine DHI milk recording, this trait would be available for every test-date, multiple times during the lactation.

The heritabilities of the predicted MFG size traits were both less than those found for their corresponding gold standard measured trait (0.344 and 0.301 for pD[4,3] and pD[3,2],
respectively). The decrease in heritability was more marked between D[3,2] and pD[3,2] than between D[4,3] and pD[4,3], but this difference was not statistically significant. It is conceivable that more error is introduced into the predicted trait due to the poor performance of the MIR calibration model in the tails of the MFG size distributions, thereby resulting in lower heritabilities. All of the variance components associated with predicted MFG size were lower than the same components for measured MFG size (Table 6.3). This is attributable to a loss in the total phenotypic variability in the predicted values caused by the prediction models. The relative decline was greater for the additive genetic variance component compared to the permanent environment and residual components, resulting in the reduced heritability estimates. The permanent environment variance remained nearly the same between D[3,2] and pD[3,2], and the decline was observed in additive genetic and residual variance components. Cecchinato et al. (2009) observed the opposite for measured and MIR predicted rennet coagulation time and curd firmness, finding that MIR predicted heritabilities were greater than the heritabilities estimated for their measured counterparts. A larger sample size would be required to properly evaluate if there are true differences in the heritabilities of measured and predicted MFG size.

**Genetic Correlations**

Genetic correlations between the examined traits are shown in Table 6.4. There was a very strong genetic correlation (0.975) found between D[4,3] and D[3,2]. These two measures of average MFG are heavily related as they are calculated from the same distribution but are biased in different directions. The surface moment mean (D[3,2]) is more sensitive to the presence of small MFG while D[4,3] is more sensitive to large particles in the globule size distribution. Therefore, a genetic correlation close to 1 was expected. The genetic correlation between D[4,3] and fat percentage was 0.536 while the genetic correlation between D[3,2] and fat percentage was 0.693. These genetic correlations may again be explained in part by the relationship between MFG size and DGAT1 genotype found by Argov-Argaman et al. (2013). A significant phenotypic relationship between average MFG size and milk fat percentage was previously described for this population by Fleming et al. (2016b), in which they reported Pearson correlation coefficients of 0.34 and 0.36 for D[4,3] and D[3,2], respectively. A phenotypic relationship between fat production and MFG has also been noted by Wiking et al. (2004) in Danish Holstein cows.
A strong genetic correlation (0.721) between measured D[4,3] and its predicted equivalent, pD[4,3] was found. The genetic correlation between D[3,2] and pD[3,2] was marginally higher at 0.792. The greater genetic correlation for D[3,2] measures may be due to the slightly greater $R^2_{cv}$ of the predictive model compared to that for D[4,3]. If the predictive models perform with greater accuracy, true and predicted values would be closer to unity and genetic correlation would seemingly be greater. Cecchinato et al. (2009) presented similar genetic correlations ranging from 0.71 to 0.87 between measured and MIR predicted curd firmness from prediction models having $R^2$ values of 0.46 to 0.52. They also noted genetic correlations ranging from 0.91 to 0.96 between measured and predicted rennet coagulation time, which were produced from models with higher $R^2$ values of 0.61 to 0.69. Due to the high genetic correlation between the measured and MIR predicted traits they found that using MIR predictions as indicator traits would have equal to slightly lower response compared to direct measurement. Although the utilized MIR prediction equations for MFG size are not precise enough to accurately quantify MFG, the strong genetic correlation suggests that selecting animals based on the predicted values could still have realized effects on true MFG size.

The critical barrier to using the MIR predicted MFG size for selection is revealed in the genetic correlation with fat percentage. The correlation of fat percentage with pD[4,3] and with pD[3,2] were 0.978 and 0.972, respectively, which is notably greater than what was observed for D[4,3] and D[3,2] (Table 6.4). The increase in the genetic correlation between fat percentage and the predicted MFG size phenotypes is likely an outcome of the prediction equation. Fat content is dominant in the milk MIR spectra and since there is a seemingly weak relationship between the spectra and average MFG size, the prediction equation is utilizing the indirect correlation with fat content. Eskildsen et al. (2014) advised that phenotypic MIR predictions of individual fatty acids may be very much connected to total fat and thus if the correlation between fatty acid contents and total fat is inconsistent, the prediction of future samples may be less accurate. This may be exceedingly true for the MFG size MIR predictions. The same phenomenon is again seen, but to a lesser extent, with protein percentage. The genetic correlation with protein percentage increased to 0.830 with pD[4,3] and 0.803 with pD[3,2]. Protein also is a significant component of the milk MIR spectra and there may be some phenotypic correlation that is being augmented through the prediction equation. Therefore, despite the strong genetic correlation between the measured and predicted MFG size, the predicted values are more genetically
associated with fat percentage and selection based on the predicted MFG size would be more so putting selection pressure on fat percentage. Additional work can be done with a larger population having predicted MFG to determine more precisely the genetic correlation between the predicted values and production traits.

As a result of the small number of cows and limited samples there are large standard errors associated with most of the estimated genetic parameters making definitive conclusions difficult. The determination of average MFG size using gold standard integrated light scattering laboratory methods is not feasible for a large number of samples on an ongoing basis like would be required for routine genetic evaluations. Mid-infrared predicted traits are attractive for selection of animals for difficult to phenotype traits and potentially valuable traits such as MFG size. Before the real benefit of a MIR predicted trait in genetic evaluations can be completely understood, it is important to have knowledge of the similarity of the genetic relationships between measured values, predicted values, and other correlated traits of interest. A comparison of the genetic parameters for gold standard measured traits and their MIR predicted equivalents could frequently be restricted by a limited number of gold standard measured records. This may result in genetic parameter estimates with large standard errors, but differences may still be distinguished and tendencies noted. These comparisons could reveal additional details of the information the prediction models are employing and if they would yield different results in correlated traits than their true values.

CONCLUSIONS

Milk fat globule size is a moderately heritable trait and it may therefore be possible to alter the average size through selection. Genetic evaluation of this trait is limited by MFG size being arduous to phenotype. MIR predicted MFG size, which could be obtained during routine milk recording, could be used as an indicator trait in genetic selection due to the high genetic correlation between measured and MIR predicted average MFG size. However, the very strong genetic correlation that appears between predicted MFG size and fat percentage, which was not as strong between measured MFG size and fat percentage, indicates that there is little utility of MIR predicted MFG size in selection programs.
Table 6.1. Descriptive statistics of milk fat globule volume moment mean (D[4,3]), surface moment mean (D[3,2]), predicted volume moment mean (pD[4,3]), predicted surface moment mean (pD[3,2]), and milk fat and protein percentages.

<table>
<thead>
<tr>
<th>Trait</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>D[4,3] (µm)</td>
<td>1,583</td>
<td>4.130</td>
<td>0.483</td>
<td>2.78</td>
<td>6.09</td>
</tr>
<tr>
<td>D[3,2] (µm)</td>
<td>1,586</td>
<td>3.468</td>
<td>0.339</td>
<td>2.21</td>
<td>4.73</td>
</tr>
<tr>
<td>pD[4,3] (µm)</td>
<td>1,551</td>
<td>4.171</td>
<td>0.357</td>
<td>3.17</td>
<td>5.67</td>
</tr>
<tr>
<td>pD[3,2] (µm)</td>
<td>1,549</td>
<td>3.489</td>
<td>0.243</td>
<td>2.75</td>
<td>4.36</td>
</tr>
<tr>
<td>Fat (%)</td>
<td>3,643</td>
<td>3.951</td>
<td>0.704</td>
<td>1.34</td>
<td>8.04</td>
</tr>
<tr>
<td>Protein (%)</td>
<td>3,643</td>
<td>3.282</td>
<td>0.357</td>
<td>2.42</td>
<td>4.78</td>
</tr>
</tbody>
</table>

Table 6.2. Estimated heritabilities (h^2), their associated standard errors (SE), and repeatabilities (r) of milk fat globule volume moment mean (D[4,3]), surface moment mean (D[3,2]), predicted volume moment mean (pD[4,3]), predicted surface moment mean (pD[3,2]), and milk fat and protein percentages.

<table>
<thead>
<tr>
<th>Trait</th>
<th>h^2</th>
<th>SE</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>D[4,3]</td>
<td>0.392</td>
<td>0.183</td>
<td>0.579</td>
</tr>
<tr>
<td>D[3,2]</td>
<td>0.494</td>
<td>0.185</td>
<td>0.623</td>
</tr>
<tr>
<td>pD[4,3]</td>
<td>0.344</td>
<td>0.189</td>
<td>0.550</td>
</tr>
<tr>
<td>pD[3,2]</td>
<td>0.301</td>
<td>0.192</td>
<td>0.559</td>
</tr>
<tr>
<td>Fat (%)</td>
<td>0.426</td>
<td>0.157</td>
<td>0.432</td>
</tr>
<tr>
<td>Protein (%)</td>
<td>0.299</td>
<td>0.170</td>
<td>0.503</td>
</tr>
</tbody>
</table>
Table 6.3. Estimated additive genetic ($\sigma_a^2$), permanent environment ($\sigma_{pe}^2$), residual ($\sigma_e^2$), and total phenotypic ($\sigma_p^2$) variances of milk fat globule volume moment mean (D[4,3]), surface moment mean (D[3,2]), predicted volume moment mean (pD[4,3]), predicted surface moment mean (pD[3,2]), and milk fat and protein percentages.

<table>
<thead>
<tr>
<th>Trait</th>
<th>$\sigma_a^2$</th>
<th>$\sigma_{pe}^2$</th>
<th>$\sigma_e^2$</th>
<th>$\sigma_p^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>D[4,3]</td>
<td>0.068</td>
<td>0.032</td>
<td>0.073</td>
<td>0.172</td>
</tr>
<tr>
<td>D[3,2]</td>
<td>0.043</td>
<td>0.011</td>
<td>0.033</td>
<td>0.088</td>
</tr>
<tr>
<td>pD[4,3]</td>
<td>0.031</td>
<td>0.018</td>
<td>0.040</td>
<td>0.089</td>
</tr>
<tr>
<td>pD[3,2]</td>
<td>0.013</td>
<td>0.011</td>
<td>0.019</td>
<td>0.043</td>
</tr>
<tr>
<td>Fat (%)</td>
<td>0.163</td>
<td>0.002</td>
<td>0.218</td>
<td>0.384</td>
</tr>
<tr>
<td>Protein (%)</td>
<td>0.019</td>
<td>0.013</td>
<td>0.031</td>
<td>0.063</td>
</tr>
</tbody>
</table>

Table 6.4. Genetic correlations with standard errors in parentheses of milk fat globule volume moment mean (D[4,3]), surface moment mean (D[3,2]), predicted volume moment mean (pD[4,3]), predicted surface moment mean (pD[3,2]), and milk fat and protein percentages.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>D[4,3]</td>
<td>0.975 (0.027)</td>
<td>0.721 (0.194)</td>
<td>0.700 (0.220)</td>
<td>0.537 (0.228)</td>
<td>0.462 (0.358)</td>
</tr>
<tr>
<td>D[3,2]</td>
<td>0.829 (0.149)</td>
<td>0.792 (0.151)</td>
<td>0.692 (0.153)</td>
<td>0.733 (0.336)</td>
<td></td>
</tr>
<tr>
<td>pD[4,3]</td>
<td>0.969 (0.028)</td>
<td>0.978 (0.147)</td>
<td>0.830 (0.328)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pD[3,2]</td>
<td>0.972 (0.123)</td>
<td>0.803 (0.344)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Chapter 7
GENERAL DISCUSSION AND CONCLUSIONS

Mid-infrared (MIR) spectroscopy provides an opportunity for the dairy industry beyond its current role in milk recording programs. This thesis developed prediction equations for fatty acid (FA) contents and average milk fat globule (MFG) size from milk samples with gold standard measurements and MIR spectra collected during routine DHI milk recording, and evaluated their utility for genetic improvement in Canada.

MILK FAT GLOBULE SIZE

Variation in average MFG size was observed and this was related to various factors including the genetics of the cow. A moderate heritability was found for average MFG size in milk suggesting that there is a possibility of selecting animals for this trait for the improvement of the nutritional value or technological properties of milk. However, access to large numbers of phenotypes is needed. The ability to predict average MFG size from MIR spectra was found to be poor, although there was a relatively strong genetic correlation between the gold standard measured and MIR predicted MFG size values. The primary obstacle in using MIR predicted MFG size is its strong phenotypic and genetic correlation with milk fat content. These correlations were greater than the correlations found between the measured MFG size and fat content. Thus, the prediction equation is heavily utilizing the relationship between MFG size and total fat content, which dominates the spectrum. The very strong genetic correlation between the predicted MFG size and fat content concedes that selection using the predicted values would essentially be the same as selecting for fat content. However, the estimation of genetic parameters was hampered by small sample sizes and the estimates had large standard errors associated with them. The developed MIR predictions of MFG size should not be used for selection purposes, because of their low accuracy and high correlations with fat content. The utility of the MFG prediction equation seems to be limited to screening milk samples for either small or large globules on average. This could aid in directing milk with more favorable MFG size toward the production of milk products affected more by MFG size, such as cheese.
MILK FATTY ACIDS

More success was observed for the prediction of FA contents in milk from MIR spectra. The ability to predict FA contents varied and was largely dependent on the concentration of the FA in milk. The defined groups of FA and those individual FA making up a high proportion of the total fat had high prediction accuracies, while those FA with minor contributions predicted poorly. The calibration set used to create the prediction equation was shown to be extremely critical and can have a great effect on the outcome. Significant variation needs to be contained in the calibration set, and the variation expected in the entire population for which the prediction is to be used must be well represented. Improved results were found when the calibration set had a more uniform distribution, created by removing the surplus of samples with FA quantities close to the mean. The collection of additional FA with more extreme FA values may assist in predicting extreme values with more accuracy. A critical next step will be the validation of the FA prediction equations using an external population. The accuracy of the predictions needs to be evaluated on milk samples of Canadian cows distant from the samples used to build the prediction equations. This process will give a better indication on how well the predictions will work for the entire Canadian dairy cow population. In the future, additional novel methodologies for producing the prediction equations can also be explored in order to improve the model accuracies.

Groups of FA and the major individual FA in milk can now be predicted with sufficient accuracy using the developed prediction equations. These equations can be used on the historical, saved spectra and for all incoming milk samples. It is the utility of this new information that still needs to be addressed.

FUTURE OPPORTUNITIES

There are many potential applications for the MIR predicted FA contents seeing that these phenotypes can be available for all test-day milk samples for every cow enrolled in Canadian DHI milk recording programs. One possibility that could be examined is their use as a management tool. Deviations from the expected FA contents in milk for a cow could be a signal that intervention is required. The access to these phenotypes will allow for more research into different managerial employments that could include feed and energy intakes, energy status, reproduction considerations, and udder and systemic health events.
The predicted FA phenotypes also provide a considerable opportunity for dairy cattle genetic improvement programs. The groups of FA based on chain-length and saturation were found to be moderately heritable, and genetic variation exists in the Canadian population so selection for these traits could proceed. One obstacle is the defining of a breeding goal. The perceived optimal FA composition can vary among human nutritionists and dairy technologists. Thus, there may be a different target composition for milk with different destinations and changes in fat composition could potentially improve one milk product but harm another. As well, because of the link between milk composition and the cow physiological status, as discussed by Hamann and Krömker (1997), there may be concerns for the animal that need to be understood before a FA composition goal can be defined.

Prior to the incorporation of FA component traits into breeding objectives, it is crucial that their full genetic influence is known in order to avoid negative correlated responses. The five fatty acid groups explored had similar genetic correlations with the production traits, and these were similar to those for fat percentage. However, medium-chain and saturated fatty acid groups were found to have stronger average daily genetic correlations with fat yield, fat and protein percentage, and fat to protein ratio compared to those found for long-chain and unsaturated fatty acid groups. Therefore, selection focused on these production traits may increase medium-chain and saturated fatty acid contents more than the others. The present work also noted that the correlations between fatty acids and the fat related productions traits were not constant across the lactation, and it was in the beginning stage of lactation where differences were most evident. At this stage of lactation, selection for increased total fat could have a greater impact on long-chain and unsaturated fatty acids. Focusing on increasing long-chain fatty acids in early lactation, however, could mean increased problems related to the effects of negative energy balance for the cow. There may be an optimal lactation stage to define the FA traits at, if selection for FA composition changes were to be performed, so that any potential negative ramifications for the cow are minimized. In the future this work should be expanded to focus on the genetic correlations among FA composition and cow health and fertility traits recorded in Canada. According to Bastin et al. (2016), traits predicted using MIR spectroscopy, including FA, are good candidates as indicator traits in dairy breeding programs for health and fertility. This would most likely be the first application of FA in breeding programs in Canada because of the lack of a clear breeding goal for FA composition and no current incentives for implementing change.
LIMITATIONS

Predicted phenotypes from MIR spectra are not the true values of the actual trait, and their precision depends on the accuracy of the developed prediction equations. The predictive models may not work across all populations, and similar variation needs to be in the calibration set and population to be predicted. If the variation and range of a trait changes in a population, previously developed equations may no longer predict the trait sufficiently. Therefore, validation of the models in each population is vital to knowing their utility. As well, although MIR spectrometers use the same technology, their spectra is subject to variations between instruments and over time. This can affect the accuracy of predictions from the spectra produced and regular standardization of spectrometers is required.

One potential unknown for MIR calibration equations are how robust these models will be in the long term for various traits. Currently used, and extremely accurate, equations for total fat and protein contents are successful because there is a clear relationship between these traits and MIR spectra. In particular, the alkyl, carbonyl, and amide chemical groups of fats and proteins have strong absorption bands in the MIR region and allow for their direct quantification. Much like was observed in the prediction of MFG size, for some traits with less of a direct correlation with the MIR spectrum, models are using information from other correlated components to predict the trait of concern. This idea was also presented using milk samples spiked with acetone (Heuer et al., 2001) and individual fatty acids (Eskildsen et al., 2014). Prediction equations based more so on these indirect correlations only hold as long as these correlations remain stable (Eskildsen et al., 2014). In more complex traits for which MIR predictions do not have an explicit connection to the spectra (e.g. energy balance, methane emissions), it will become more important to understand what components of the spectra are being exploited and whether the predictions will be robust. Care needs to be taken and the predictive ability should be regularly confirmed or adjusted for all calibration equations to ensure their longevity.

CONCLUSIONS

There is great potential in the use of mid-infrared spectroscopy for the prediction of novel trait phenotypes, including fatty acids, in the Canadian dairy industry. The integration of these traits into management and breeding practices may serve to improve milk quality and
technological aspects, cow health and fertility, and environmental concerns. Presently there is an opportunity to implement fatty acid predictions into milk recording programs and use these phenotypes for the genetic improvement of milk fat composition or possibly increase the accuracy of selection for other important traits. Although this methodology is still in its infancy and there are many considerations to be made, MIR predicted fatty acids can be produced and provide added-value to the industry.
REFERENCES


Selection for milk coagulation properties predicted by Fourier transform infrared spectroscopy in the Italian Holstein-Friesian breed. J. Dairy Sci. 97:4512-4521.


Ferrand, M., B. Huquet, S. Barbey, F. Barillet, F. Faucon, H. Larroque, O. Leray, J. M.


Hansen, P. W. 1999. Screening of dairy cows for ketosis by use of infrared spectroscopy and


Rutten, M. J. M., H. Bovenhuis, and J. A. M. van Arendonk. 2010. The effect of the number of observations used for Fourier transform infrared model calibration for bovine milk fat


