Energy Balance and Metabolic Status of Dairy Cows

A study using metabolomics, proteomics and machine learning approaches

Wei Xu

Energy balance and metabolic status of dairy cows

a study using metabolomics, proteomics and machine learning approaches

Wei Xu

Thesis committee

Promotor

Prof. Dr B. Kemp
Professor of Adaptation Physiology
Wageningen University & Research

Co-promotors

Dr A.T.M. van Knegsel
Senior Researcher, Adaptation Physiology Group
Wageningen University & Research

Dr J.J.M. Vervoort
Associate Professor, Laboratory of Biochemistry
Wageningen University & Research

Other members

Prof. Dr R.F. Veerkamp, Wageningen University & Research
Prof. Dr K. Huber, University of Hohenheim, Germany
Dr M.M. Hostens, Utrecht University
Dr R. Stienstra, Wageningen University & Research

This research was conducted under the auspices of the Graduate School of Wageningen Institute of Animal Sciences (WIAS).

Energy balance and metabolic status of dairy cows

a study using metabolomics, proteomics and machine learning approaches

Wei Xu

Thesis
submitted in fulfilment of the requirements of the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus
Prof. Dr A. P. J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Thursday 12 December 2019
at 11.00 a.m. in the Aula.

Wei Xu

Energy balance and metabolic status of dairy cows, a study using metabolomics, proteomics and machine learning approaches

163 Pages

PhD thesis, Wageningen University, Wageningen, the Netherlands (2019)

With references, with summary in English

ISBN: 978-94-6395-130-2

DOI: 10.18174/500864

Table of Contents

CHAPTER 1

General Introduction

1.1. Introduction

Over the past decades, milk yield of modern high-producing dairy cows has reached to on average 9,000 kg per lactation due to genetic selection, nutritional strategies, and improved farm management [1,2]. In early lactation, milk yield typically peaks in approximately lactation week 7 [3,4], however, the maximum feed intake is reached between lactation week 9 and 15 [5,6]. As a result, dairy cows usually experience a negative energy balance (NEB) in early lactation due to the rapid increased energy requirements for milk production and limited energy intake from feed [7,8]. Negative energy balance and related metabolic status are associated with a greater risk of metabolic disorders [9], disordered fat metabolism in the liver [10-12], infectious diseases [8], reduced fertility [13,14], immune response of mammary gland [15,16], and a short productive lifespan of dairy cows [17].

On farm, early and reliable screening of energy balance and metabolic status could identify dairy cows with an increased risk for health and fertility problems. Currently, energy balance could be estimated by milk characteristics, such as milk yield, fat, protein, lactose, days in milk [18], and body characteristics, such as, body weight change and body condition score (BCS) of dairy cows [19]. In addition, those variables combined with ketone bodies in plasma or milk are also used to indicate metabolic status of dairy cows in early lactation [20-22]. The current screening methods, however, have limitations in practice. For example, milk characteristics have good capability to estimate energy balance of cows at herd level ($R^2 = 0.90$, proportion of the variance in the energy balance that is estimated from the milk characteristics), but a limited capability to estimate the energy balance of individual cows ($R^2 = 0.40$) [21,23]. Although plasma biomarkers are more precise to indicate metabolic status than milk characteristics [24], blood collection is invasive and labour intensive [25,26]. An alternative method is to estimate energy balance and metabolic status of dairy cows with on-farm cow data or milk characteristics that are easily obtained in practice. Moreover, a clear metabolic pathway related to energy balance and the alteration of metabolic status helps to understand health and metabolism of dairy cows during NEB. Understanding metabolism of dairy cows during NEB could facilitate development of strategies to manage health and metabolism of dairy cows in early lactation.

This chapter will first describe the NEB of dairy cows in early lactation, and how the NEB is related to altered metabolic status. Secondly, knowledge on indicators for energy balance and metabolic status in dairy cows is reviewed. Subsequently, metabolomics, proteomics, and machine learning techniques are briefly introduced as well as their applications in dairy cows. Lastly, the aim and outline of this thesis is presented.

1.2. Negative Energy Balance and Metabolic Status of Dairy Cows in Early Lactation

In dairy cows, elevated energy requirements for milk production combined with a relatively low energy intake from feed result in a NEB in early lactation [27,28]. During NEB, dairy cows mobilize body reserves to meet the energy requirement for maintenance and milk production (Fig. 1.1). Mobilized body reserves includes mainly body fat (74% to 93%) and for a small proportion also body protein (6% to 7%) [29,30]. Subsequently, body reserve mobilization is associated with an altered concentration of plasma metabolites and metabolic hormones, such as, increased free fatty acids (FFA), β-hydroxybutyrate (BHB), and growth hormone; and decreased glucose, insulin, and insulin-like growth factor 1 (IGF-1) (Fig. 1.2). Therefore, a NEB status of dairy cows is related to an altered metabolic status in early lactation, including carbohydrate, lipid, and protein metabolism, and is associated with an altered hormonal regulation.

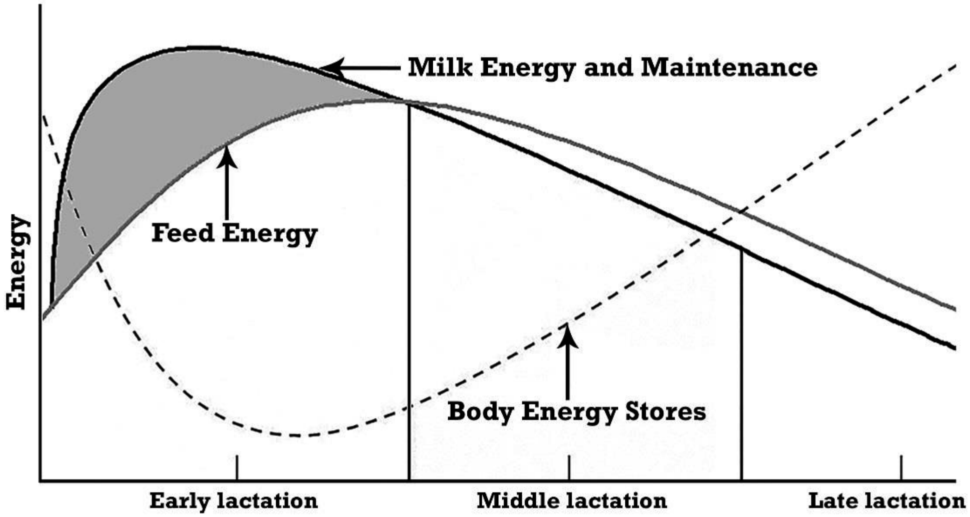


Fig. 1.1. Energy requirements for milk yield and body maintenance, energy from feed intake, and body energy stores of dairy cows during lactation [31].

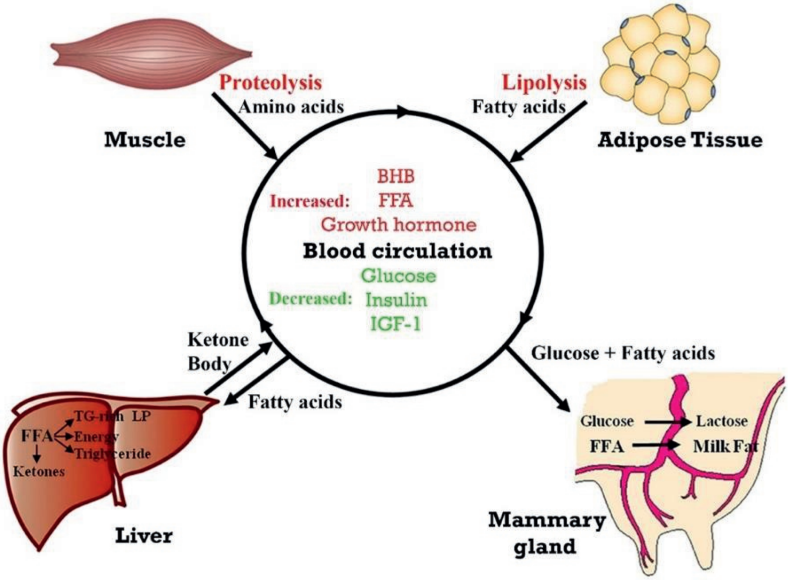


Fig. 1.2. Body reserve mobilisation and changes in metabolic status during negative energy balance of dairy cows in early lactation.

Abbreviations: BHB, β-hydroxybutyrate; FFA, free fatty acids; IGF-1, insulin-like growth factor 1.

1.3. Estimation of Energy Balance and Metabolic Status

In early lactation, precise estimation of energy balance and metabolic status of dairy cows could facilitate in optimization of dairy cow management and improve animal health and welfare. Measuring energy balance using indirect calorimetry based on in respiration chambers is considered as the golden standard approach to determine energy balance of dairy cows [32], but has limited applications even under experimental conditions due to high requirements for equipment, and budget and limitations concerning animal numbers which can be included in the experiment. Alternatively, energy balance can be estimated from either the difference between net energy intake and net energy requirement for maintenance and milk production [33], or the change in body condition, or the change in body weight [34]. However, both methods have limitations in practice. For example, input-output measure is hard to apply on commercial farms due to limited or no possibilities to measure individual energy intake; and body condition measure is not well studied to assess short-term change in energy status due to the relative insensitivity of body condition score. An alternative approach is to estimate energy balance with data from milk characteristics and farm management (Table 1.1). In these previous studies on estimating energy balance with data from milk characteristics and farm management, 5 out of 8 studies estimated the energy balance of cows at a herd level; 4 out of 8 studies estimated the energy balance of individual cows. The performance to estimate energy balance at herd level is generally higher than the performance at individual level. At herd level, up to 94.0% variation of energy balance could be explained [18], or up to 0.80 correlation was found between estimated energy balance and calculated energy balance [35]. At individual level, up to 40.0% variation of energy balance (body condition measure) and up to 50.0% variation of energy balance (input-output measure) could be explained by milk yield, fat, protein, lactose, and days in milk [18]. Up to 75.0% of variation of energy balance could be estimated by full range mid-infrared spectrometry of milk samples [36], which is the best performance of an indicator for energy balance so far. Nevertheless, mid-infrared spectrometry will not reveal alterations in metabolic pathways related to the energy balance.

Table 1.1. Studies on indicators for energy balance of dairy cows in early lactation.

Author, year (number of cows)	Lactation (wk/d)	Estimator/indicator and data sources
Grieve et al., 1986 [37] (236)	5-12 wk	Milk fat and protein
Heuer et al., 2000 [38] (72)	2-12 wk	Lactation week, parity, milk yield, fat and protein
Heuer et al., 2001 [35] (264)	2-12 wk	Lactation week, parity, milk yield, fat, protein and body weight
Reist et al., 2002 [39] (90)	1-10 wk	Lactation week, milk yield, fat, protein, lactose, and milk acetone
Friggens et al., 2007 [18] (299)	*NA*	Milk yield, fat, protein, lactose, days in milk
Mäntysaari et al., 2010 [19] (146)	0-305 d	Milk fat, body weight, and BCS.
McParland et al., 2011 [36] (288)	5-305 d	Mid-infrared spectrometry
Alphonsus et al., 2015 [40] (60)	4-305 d	Milk yield, fat, protein and lactose
Stoop et al., 2009 [41] (1933)	63-282 d	Milk fat C16:0, C18:0 ↑, odd chain C5:0 to C15:0 ↓
Gross et al., 2011 [42] (30)	1-12 wk	Milk short- and medium-chain FFA (up to C16) ↓ Milk long-chain FFA, especially C18:1 cis-9 ↑
Ducháček et al., 2013 [43] (27)	7-119 d	Milk fat, C14:0, C16:0, C18:0, C18:1, C18:2, and C16:1 ↑

Abbreviations: BCS, body condition score; FFA, free fatty acids; NA, Not available.

Besides studies to estimate energy balance, 3 studies have reported the relationship between negative energy balance and altered milk fat profiles, including decreased proportion of short-chain fatty acids, and increased proportion of medium- and long-chain fatty acids in milk fat [41-43]. The different relationship between energy balance and milk fat profiles among those studies could be explained by methods to estimate energy balance, either by changed BCS [43], or by milk fat-to-protein ratio [41], or by different between energy intake and energy expenditure [42].

Compared with studies to estimate energy balance, more studies have used plasma or milk metabolites to indicate the metabolic status of dairy cows in early lactation (Table 1.2). In total 22 studies have established indicators of subclinical ketosis or hyperketonemia, including plasma BHB (>1.0 m*M* to >1.4 m*M*), FFA (>0.26 m*M*), and milk BHB (>0.07 m*M*), acetone (>0.15 m*M*), and urea BHB (>0.14 m*M*). The variation in cut-off values, for example plasma BHB as an indicator for hyperketonemia ranged from 1.0 m*M* to 1.4 m*M* among studies, is possibly due to difference in animal breed [44], lactation stage [42], and parity [45]. Clinical ketosis is generally defined as clinical signs combined with a much higher concentration of plasma BHB (>2.5 m*M*) [46,47]. Other reported indicators for clinical ketosis are milk fat to protein ratio [48], concentration of acetone [44] milk acetoacetate [49] or feeding behaviour-relate variables [50]. Metabolic status related to hepatic lipidosis were indicated by plasma BHB (>0.78 m*M*), FFA (>0.57 m*M*), and liver enzyme aspartate aminotransferase (>120 U/L). Moreover, metabolic status of dairy cows with limited reports is not shown in the Table 1.2. For example, protein metabolism of dairy cows is indicated by the urea concentration in blood or milk [51-53].

Currently, widely used indicators for metabolic status include blood BHB, FFA, and milk BHB. These indicators were established with high correlation between metabolites and metabolic status, such as, subclinical ketosis and hepatic lipidosis. Those metabolites, however, have not been used to estimate energy balance of dairy cows. A model to estimate energy balance is expected to include those metabolites, which are related to energy metabolism of cows. Over the past decade, increasing low-abundant metabolites in blood and milk have been detected through metabolomics technique to reveal the metabolic profiles of dairy cows, which have been correlated with subclinical ketosis [54-56], heat stress [57], hepatic dysfunction [58], and lipidosis [59]. To our knowledge, however, no study has estimated energy balance of individual dairy cow through plasma or milk metabolomics data.

1.4. New Potential Approaches to Estimate Energy Balance and Metabolic Status

In previous studies, plasma metabolites and hormones were used as an indicator for metabolic status [60], but blood sampling is invasive and labour intensive. Therefore, a non-invasive approach to estimate energy balance or metabolic status using milk samples or on-farm cow data could be interesting. In this thesis, we hypothesize that energy balance and metabolic status could be estimated by *i*), milk characteristics combined with detected milk metabolites by metabolomics approach in reduced models; and *ii*), on-farm cow data, including milk characteristics, based on machine learning techniques.

1.4.1. Metabolomics and proteomics

Metabolomics is a high-throughput technique that can identify, quantify and characterize hundreds to thousands of metabolites from bio-samples using targeted or global analytical approaches [61,62]. Compared with studies that focus on one or two metabolic pathways, metabolomics studies can describe a broad range of metabolites involved in different metabolic pathways. In dairy cows, milk metabolomics profiles described by metabolomics studies have been related to metabolic disorders, diseases, and productive lifespan in dairy cows [57,63]. Milk phosphocholine and phosphatidylcholine were related to ketosis [64] and heat stress [57]. Compared with healthy cows, less carnitine, citrate, and hippurate were detected in milk of cows with clinical mastitis [63]. In addition, milk metabolites would be expected to be associated with metabolism in the synthesis of milk production. Milk citrate is involved in milk fat synthesis [65], and milk phosphate sugars are intermediates in milk lactose synthesis [66]. To our knowledge, however, metabolomics techniques have not been applied to estimate energy balance of individual cows. We hypothesize that milk metabolites detected through metabolomics techniques can estimate energy balance of individual cows in early lactation.

Proteomics techniques aim to systematically detect, quantify, and characterize proteins from biological samples [67]. In bovine milk, more than 90% of protein composes of three major proteins, α_{s1}-casein (30%), β-casein (30%), and β-lactoglobulin (30%) [68]. Several hundreds of low-abundant proteins in milk are in addition present [69]. These low-abundant proteins make milk proteomics studies complicated, especially when combined with post-translational modifications [70]. Despite this fact, milk proteomics can be used to study changes in metabolic pathways of dairy cows through the expression or modification of proteins [71,72]. Through milk proteomics studies, researchers have investigated the metabolic pathways of lipid synthesis and protein secretion in the mammary gland of dairy cows [71], and identified biomarkers of mastitis due to the association between milk proteins and intra-mammary host defence mechanisms [73,74]. To our knowledge, however, no proteomics study has been executed to reveal the metabolic pathway related to energy metabolism in the mammary gland of dairy cows. It can be hypothesized that milk proteomics data could not only confirm our findings of milk metabolomics studies, but also reveal the metabolic pathways related to energy balance of dairy cows through the altered protein expression or modification.

1.4.2. Machine learning techniques

With the increased calculating capability of computers, the analysis of large and complex datasets could be realized through various flexible algorithms using machine learning techniques. Machine learning techniques can estimate cow performance or disease events using statistical methods to handle datasets with high dimensional variables [75,76]. In early lactation, energy expenditure of dairy cows is associated with milk yield components, parity, and BCS [77,78], which are available data on most dairy farms. Therefore, those on-farm cow data can be expected to estimate metabolic status of cows. Available on-farm cow data, however, usually include a series of correlated variables, for example, fat yield and milk yield are correlated [79]. Therefore, it is difficult to define a specific function (*e.g.*, linear, quadratic, etc.) with conventional statistical methods. Machine learning techniques can deal with complicated correlations caused by ever increasing number of variables [80]. During the past decade, machine learning techniques were used to estimate milk yield [81], breeding values [75], calving time [82], reproductive performance [76,83], and even to identify mastitis [84,85] in dairy cows. It can be hypothesized that on-farm cow data can estimate energy balance and metabolic status of dairy cows using machine learning techniques.

1.5. Aim and Outline of this Thesis

The aims of this thesis are, first to estimate energy balance and metabolic status of dairy cows using metabolomics and machine learning techniques; and second to investigate the metabolic pathway related to energy metabolism of dairy cows in early lactation using metabolomics and proteomics techniques.

To study the first aim, data were collected from two earlier studies: study I [86], and study II [87]. In study I, 168 cows were assigned randomly to 3 groups with 0-d, 30-d or 60-d dry period length (DPL). After one lactation within the experiment, 130 cows were monitored for a second lactation within study I [88]. In study II, 127 cows were assigned randomly to 2 groups with either a 0-d or 30-d DPL. To study the second aim, milk and blood samples obtained from study II are analysed in a metabolomics and proteomics study. The specific aims and methods in each chapter in this thesis are briefly introduced here.

Chapter 2 describes a metabolomics approach to estimate the energy balance of dairy cow using milk characteristics and milk metabolites detected by liquid chromatography mass spectrometry (LC-MS). In addition, the metabolic pathways related to energy metabolism are discussed.

Chapter 3 describes a metabolomics approach to reveal more complex metabolic pathways related to energy balance in mammary gland of dairy cows using milk metabolites detected either by LC-MS or nuclear magnetic resonance spectroscopy (NMR).

Chapter 4 describes a metabolomics approach to investigate the relation between milk metabolites and plasma metabolites using milk and blood metabolites detected by LC-MS and NMR.

Chapter 5 describes a preliminary study on a proteomics approach related to the energy balance of cows in early lactation.

Chapter 6 describes how on-farm cow data can be used to estimate metabolic status of dairy cows in early lactation using machine learning algorithms.

Chapter 7 describes how on-farm cow data can be used to estimate hyperketonemia of dairy cows in early lactation using machine learning algorithms.

Finally, Chapter 8 is the general discussion chapter reviewing all data obtained in this thesis with metabolic pathways related to energy metabolism, the performance of machine learning algorithms in practice as well as future perspectives.

Table 1.2. Studies on indicators for metabolic status of dairy cows in early lactation.

Author, year (number of cows)	Lactation (wk/d)	Indicators or methods
Hepatic lipidosis and function		
Acorda *et al*. 1995 [89] (158)	*NA*	Digital analysis and ultrasonography
Starke et al., 2010 [90] (151)	35±5 d	Liver ultrasonographic images
Mostafavi et al., 2013 [91] (506)	*NA*	Plasma BHB >0.78 m*M*, FFA >0.57 m*M*, FFA/cholesterol ratio >0.2, AST >120 U/L
Imhasly et al., 2014 [92] (28)	1-6 wk	Plasma glycine, glutamine, PtC derivatives and SM derivatives
Mostafavi et al., 2015 [10] (506)	*NA*	Plasma fructosamine <0.22 m*M*
Xu *et al*., 2016 [59] (171)	14-21 d	Plasma BHB, glucose, and total 13 metabolites
Huber *et al*., 2016 [58] (19)	-7 to 15 wk	Plasma spermidine, carnitine, and acylcarnitines
Hyperketonemia, subclinical ketosis and clinical ketosis		
Filar, 1979 [93] (*NA*)	*NA*	CK: plasma BHB >1.00 m*M*
Erhardt *et al*., 1982 [94] (*NA*)	*NA*	SCK: milk citrate
Whitaker *et al*., 1983 [95] (3000)	*NA*	SCK: plasma BHB >1.00 m*M*
Simensen *et al*., 1990 [49] (*NA*)	0-6 wk	CK: milk acetoacetate >0.10 m*M*
Nielen *et al*., 1994 [96] (185)	0-9 wk	SCK: milk BHB >1.40 m*M*; urine BHB >1.40 m*M*
Dirksen *et al*., 1995 [97] (*NA*)		Milk BHB >0.10 m*M*
Gustafsson *et al*., 1996 [44] (11690)	1-2, 5-6 wk	CK: milk acetone >1.4 m*M*
Steen *et al*., 1996 [98] (2168)	1-11 wk	SCK: milk fat/protein ratio >1.4 to >2.0
Geishauser *et al*., 1998 [99] (266)	1-3, 6, 9 wk	SCK: milk BHB >0.10 m*M*
Hünniger, 1998 [100] (*NA*)	*NA*	Milk acetone >0.25 m*M*
Hansen, 1999 [101] (310)	1-6 wk	Prognostic ketosis: milk acetone >0.27 m*M*
Dobbelaar *et al*., 1998 [102] (*NA*)	*NA*	HYK: change of milk fat
Heuer *et al*., 1999 [48] (1335)	*NA*	CK: milk fat/protein ratio >1.5
Geishauser *et al*., 2000 [103] (469)	1 wk	SCK: milk acetoacetate >0.10 m*M*
Enjalbert *et al*., 2001 [22] (60)	*NA*	SCK: milk BHB >0.07 m*M*
Carrier *et al*., 2004 [104] (2500)	2-15 d	SCK: milk BHB >0.10 m*M*
Clark *et al*., 2005 [105] (23)	1-4 wk	SCK: milk acetone >0.14 m*M* for SCK
de Roos *et al*., 2007 [106] (1080)	*NA*	SCK: milk acetone >0.15 m*M*; milk BHB >0.10 m*M*
Van Haelst *et al*., 2008 [107] (16)	3-5 wk	SCK: milk fatty acids C18:1 cis-9 ↑
González *et al*., 2008 [50] (*NA*)	*NA*	CK: feed intake, feeding time, and feeding rate ↓
Goldhawk *et al*., 2009 [108] (101)	-3 to 3 wk	SCK: feed intake, feeding time, and visits to the feeder ↓
Asl *et al*., 2011 [109] (100)	2, 4, 6 wk	SCK: plasma FFA >0.26 m*M*
Gross *et al*., 2011 [42] (20)	1-12 wk	Milk short- and medium-chain FFA (up to C16) ↓ Milk long-chain FFA, especially C18:1 cis-9 ↑
Klein *et al*., 2012 [64] (321)	*NA*	Prognostic ketosis: milk ratio GPC/PC >2.50
Sun *et al*., 2014 [56] (81)	7-21 d	SCK: plasma choline, glutamine, and total 25 metabolites
Zhang *et al*., 2013 [55] (76)	1 d	CK: plasma citrate, glycine, and total 40 metabolites
Vanholder *et al*., 2015 [110] (1715)	7-14 d	Colostrum at first milking ↑, BCS >3.25, parity >2, high in Jan-Sept, high milk yield and fat, but low milk protein
Xu *et al*., 2015 [111] (40)	*NA*	CK: 39 peptide peaks differed, 26 identified peptides in urine
Mann *et al*., 2016 [112] (70)	1-2 wk	HYK: cis-9 C16:1, cis-9 C18:1, C6/8/10/12/14/15:0
Zhang *et al*., 2017 [113] (26)	time series	CK: plasma lysine, and 24 plasma metabolites differs before, during and after ketosis

Abbreviations: BCS, body condition score; BHB, β-hydroxybutyrate; CK, clnical ketosis; FFA, free fatty acids; HYK, hyperketonemia; GPC, glycerophosphocholine; NA, Not available; PC, phosphocholine; SCK, subclinical ketosis.

CHAPTER 2

Milk Metabolomics Data Reveal the Energy Balance of Dairy Cows in Early Lactation

Wei Xu[1, 2], Jacques Vervoort[2], Edoardo Saccenti[3], Renny van Hoeij[1], Bas Kemp[1] & Ariette van Knegsel[1]

1. Adaptation Physiology Group, Department of Animal Sciences, Wageningen University & Research, the Netherlands

2. Laboratory of Biochemistry, Wageningen University & Research, the Netherlands

3. Laboratory of Systems and Synthetic Biology, Wageningen University & Research, the Netherlands

***Scientific Reports* 8, Article number: 15828 (2018)**

2.1. Abstract

In early lactation, dairy cows typically have a negative energy balance which has been related to metabolic disorders, compromised health and fertility, and reduced productive lifespan. Assessment of the energy balance, however, is not easy on the farm. Our aims were to investigate the milk metabolic profiles of dairy cows in early lactation, and to obtain models to estimate energy balance from milk metabolomics data and milk production traits. Milk samples were collected in week 2 and 7 after calving from 31 dairy cows. For each cow, the energy balance was calculated from energy intake, milk production traits and body weight. A total of 52 milk metabolites were detected using LC-QQQ-MS. Data from different lactation weeks was analysed by partial least square (PLS), the top 15 most relevant variables from the metabolomics data related to energy balance were used to develop reduced linear models to estimate energy balance by forward selection regression. Milk fat yield, glycine, choline and carnitine were important variables to estimate energy balance (adjusted-R^2: 0.53 to 0.87, depending on the model). The relationship of these milk metabolites with energy balance is proposed to be related to their roles in cell renewal.

2.2. Introduction

In early lactation, the elevated energy requirements for milk production combined with a relatively low energy intake results in an energy deficit or negative energy balance (NEB) in dairy cows [27,28]. Lipid, glycogen and protein reserves are mobilized to compensate for this energy deficit [114]. A severe NEB is related to a greater risk of metabolic disorders [115], compromised health and fertility [116] and a reduced productive lifespan of dairy cows after calving [17]. Therefore, a reliable and early screening of dairy cows with a severe NEB can identify animals with an increased risk for health and fertility problems. Traditionally, the energy balance of dairy cows can be estimated by the difference between energy input by feed intake and energy output, based on milk production traits and body weight of the cows [18,117]. On commercial farms, however, feed intake of individual cows is not available. Moreover, calculation of energy balance requires detailed information on net energy derived from feed sources. An alternative way is to estimate energy balance from milk constituents, because a milk sample is easily obtained and milk composition can be daily monitored [35,39]. In previous studies, daily records for milk production traits, including fat yield, protein yield, and fat-to-protein ratio, had a good capability to estimate the energy status at herd level when using partial least square models (PLS), however, they had limited performance ($R^2 = 0.40$) to estimate the energy balance of individual cows [18,23].

Metabolomics studies, either untargeted or targeted depending on the type of sample, the type of instrumentation and the approach used, aim to detect and analyse small molecules from bio-fluids. The application of metabolomics techniques and multivariate analysis allows new insights into the metabolic fingerprint of individual animals [118-120]. Milk metabolites are supposed to be derived primarily from the activity of the mammary epithelial cells [121]. Although the biological processes responsible for the milk metabolites is not always completely clear, milk metabolites have been used to study the metabolism of dairy cows or mammary gland function [57,64,122-125]. Milk pyruvate concentration and lactate dehydrogenase activity in milk are suggested to be correlated with mammary infections [122,123], while acetate, butyrate and lactate are related to somatic cell count [124,125]. Milk phosphocholine and phosphatidylcholine have been related to ketosis [64] and heat stress [57] in dairy cows. So, milk metabolic profiles have been identified from dairy cows with different metabolic status. To our knowledge, however, milk metabolomics data have not been used to estimate the energy balance of individual dairy cows.

In this study, we considered milk metabolomics data, acquired using LC-QQQ-MS, from 31 dairy cows with two different dry period lengths (DPL, 0 or 30 days) and different parity (2nd or 3rd parity). Milk samples were collected from these cows in week 2 and 7 after calving to include a week when cows suffer from severe NEB and a week when cows recover from NEB, respectively. Energy balance of individual cows at week 2 and 7 was calculated based on energy intake, body weight and milk energy output. Our first aim was to investigate metabolic profiles in milk of individual dairy cows and to identify the important metabolites that estimate energy balance using partial least square (PLS) modelling. The second aim was to obtain reduced models to estimate the energy balance of individual cows by a limited number of selected milk metabolites.

2.3. Material and Methods

2.3.1. Animals and experimental design

The experimental protocol for the study was approved by the Institutional Animal Care and Use Committee of Wageningen University and was conducted at Dairy Campus research farm (WUR Livestock Research, Lelystad, the Netherlands). The experimental design was described previously [126]. Briefly, 31 high-yielding Holstein-Friesian dairy cows averaging 637.4±67.1 kg of body weight (in week 2 after calving) participated in this study. Cows were selected from two parities (2nd or 3rd parity) and randomly assigned to one of two DPL (0 or 30 days) before calving. Prepartum, cows with a 0 day DPL received a lactation ration based on grass silage and corn silage (6.4 MJ net energy for lactation (NE)/kg dry matter (DM)). Cows with a 30 days DPL received a dry cow ration based on grass silage, corn silage and wheat straw (5.4 MJ NE/kg DM). Postpartum, all cows received the same basal lactation ration as provided to lactating cows prepartum plus additional concentrates. Postpartum, concentrate supply increased stepwise with 0.3 kg/d till 8.5 kg/d on 28 DIM. Body weight, milk yield and feed intake were recorded daily. During lactation, cows were milked twice daily at ~0600 hours and ~1800 hours.

2.3.2. *Milk samples*

Milk samples for analysis of fat, protein and lactose percentage (ISO 9622, Qlip, Zutphen, the Netherlands) were collected four times per week (Tuesday afternoon, Wednesday morning, Wednesday afternoon, and Thursday morning), Milk samples were analysed as a pooled sample per cow per week and used to calculate average fat, protein and lactose yield per week. Milk samples for metabolomics studies were collected at Friday morning each week. All samples were collected and stored at -20 °C until analysis. Milk production traits were averaged per week for week 2 and 7. Four milk samples were omitted because the dairy cows were suffering from mastitis in sampling week, three cows in week 2 and one cow in week 7.

2.3.3. *Energy intake and energy balance*

Roughage and concentrate were supplied separately and daily intakes were recorded per individual cow using roughage intake control troughs (Insentec, Marknesse, the Netherlands). Energy balance was calculated per week according to the Dutch net energy evaluation (VEM) system, as the difference between net energy intake and the estimated net energy requirements for maintenance, and milk yield (1,000 VEM = 6.9 MJ of NE) [33].

2.3.4. *Mass spectrum measurement and data processing*

For quantification of metabolites, a targeted, standardized and quality controlled metabolic phenotyping was performed based on LC-QQQ-MS analysis. Milk serum was transferred to an Eppendorf tube. Lipid was removed by chloroform extraction, after centrifuge (12,000 rpm, 15 min). The aqueous solution was filtered to remove protein using a Pall 0.5 mL 10-kDa cut-off spin filter with centrifugation at 12,000 rpm for 15 min. The aqueous solution was diluted as appropriate and simultaneous analysed with a triple quadrupole mass spectrometer (Shimadzu LC-QQQ-MS; LCMS-8040) using the PFPP method as described earlier [127,128]. The sample injection volume used was 1 μL, and a single analysis took 25 minutes.

2.3.5. *Statistical analyses*

Multivariate analysis: the data obtained from LC-MS and milk production traits were first log transformed, then centered and scaled to unit variance. Principle component analysis (PCA) was performed on the processed data first for identifying outliers and observing general trends. Partial least square discriminant analysis (PLS-DA) was applied to discriminate lactation weeks Briefly, lactation week (week 2 and 7) was used as categorical variable Y, and milk yield, fat yield, protein yield, lactose yield, urea, fat- and protein- corrected milk (FPCM) as well as 52 milk metabolites were used as predictor variables X. Repeated double-cross validation was used to determine the optimal number of PLS components. Permutation test (5,000 permutations) was used asses the validity of PLS discriminant model and to avoid overfitting [129]. Partial least square (PLS) regression was used to investigate the association between energy balance and milk metabolites and milk production traits. The energy balance of dairy cows was continuous variable Y, and milk yield, fat yield, protein yield, lactose yield, urea, fat- and protein- corrected milk (FPCM) as well as 52 milk metabolites in week 2 (or week 7) were used as observable variables X. Double-crossed validation was used to determine the optimal number of PLS components. Permutation test (5,000 permutations) was used to assess the validity of PLS regression model and to avoid overfitting. Variable importance in projection (VIP) score were used to select the most contributing variable to the PLS model [130].

Mixed models: To analyse the effect of energy balance, DPL and parity, top 15 variables with the highest VIP scores in each dataset (week 2 or 7) were analysed using a Mixed Model approach, and DPL, parity, and their two-way interactions were included in the model as fixed effects. The statistical model used for milk metabolites and milk production traits was as follows:

$$M_{jk} = \mu + Energy\ balance + DPL_j + Parity_k + Interactions_{jk} + \varepsilon_{jk}$$

where M represents the observed level of milk metabolites and milk production traits. The mean is represented by μ. DPL_j represents the fixed class effect of DPL (j = 0 day, 30 days). $Parity_k$ represents the fixed class effect of parity (k = 2nd parity, 3rd parity). $Interactions_{jk}$ are presented as significant effect from Energy balance × DPL_j + Energy balance × $Parity_k$ + DPL_j × $Parity_k$, + Energy balance × DPL_j × $Parity_k$. Non-significant interactions were excluded from the model via backward stepwise elimination if *P-value* was more than 0.10.

Stepwise regression: For obtaining reduced models with a maximum of 4 variables to estimate energy balance, top 14 variables from milk in week 2 (or week 7) with highest VIP scores in the PLS model were analysed using ten-fold crossed validation. Briefly, an *F*-test was constructed based on forward selection approach. R^2 increases with the increased estimator in the model, however, the adjusted-R^2 increases only if the new term improves the model more than would be expected by chance. The model would be selected only with higher adjusted-R^2. The statistical model used to estimate energy balance is as follows:

$$Energy\ balance_{ijk..q} = intercept_i + M_j + M_k + ... + M_q + e$$

where Energy balance $_{ijk..q}$ represents the sum of squares of estimated energy balance, intercept$_i$ represents the sum of squares of intercept, M$_{jk...q}$ represents the sum of squares of milk metabolites and milk production traits, e represents the sum of squares of error in the model.

Multivariate analysis was performed with the R environment (version 3.3.2), PLS-DA and PLS regression were performed with using the package "MixOmics" [131], double-cross validation was executed with package "chemometrics" [132]. The correlation plots of data in week 2 (or week 7) were drawn with package "corrplot". The Mixed Model was performed through PROC MIXED of SAS version 9.3 (SAS Institute, Inc., Cary, NC). The 10 fold crossed validation for obtaining reduced models was performed in R-project with package "DAAG" [133].

2.4. Results and Discussion

2.4.1. Characterization of metabolomics profiles and multivariate analysis

In the current study, 52 milk metabolites could be identified. The exact origin of milk metabolites is not clear, and they may be secreted from mammary epithelial cells, leaked from damaged somatic cells in the mammary gland or even be transferred from blood [125,134].

Combined with data from milk production traits, the metabolomics data were subjected to multivariate analysis. Principal component analysis (PCA), on the 52 milk metabolites plus 6 milk production traits showed that the metabolomics profiles can be separated by lactation week but not by dry period length or parity (Fig. 2.1-A and *Appendix* Fig. 2.1). Partial least square discriminant analysis (PLS-DA) confirmed that data could be discriminated by lactation week (Fig. 2.1-B) with a discriminant power of $Q^2 = 0.85$, (*P*-value = 2.0×10^{-4} obtained using a permutation test).

Given the data could be separated by lactation week, further analysis was performed considering data for lactation week 2 and 7 separately. Using PLS regression, energy balance was used as the dependent variable Y, and milk yield, fat yield, protein yield, lactose yield, urea, fat- and protein- corrected milk (FPCM) and 52 milk metabolites were used as predictor variables X. The capability of the PLS models to estimate the energy balance were $Q^2 = 0.72$ and $Q^2 = 0.84$, for the PLS models for week 2 and 7 respectively, and both models were statistically significant (P-value = 2.0×10^{-4} obtained using a permutation test). The variable importance in projection (VIP) scores of all variables in the first component of PLS for week 2 and 7 are shown in Fig. 2.2. Five milk production traits (milk yield, fat yield, protein yield, lactose yield and FPCM) and five milk metabolites (glycine, choline carnitine, citrulline and proline) were selected from top 15 variables with the highest VIP scores in both lactation weeks. In week 2, the five unique milk metabolites were creatinine, cystine, cytidine monophosphate (CMP), hydroxyproline and citrate. In week 7, the five unique milk metabolites were pantothenate, creatine, acetyl-choline, serine and tyrosine.

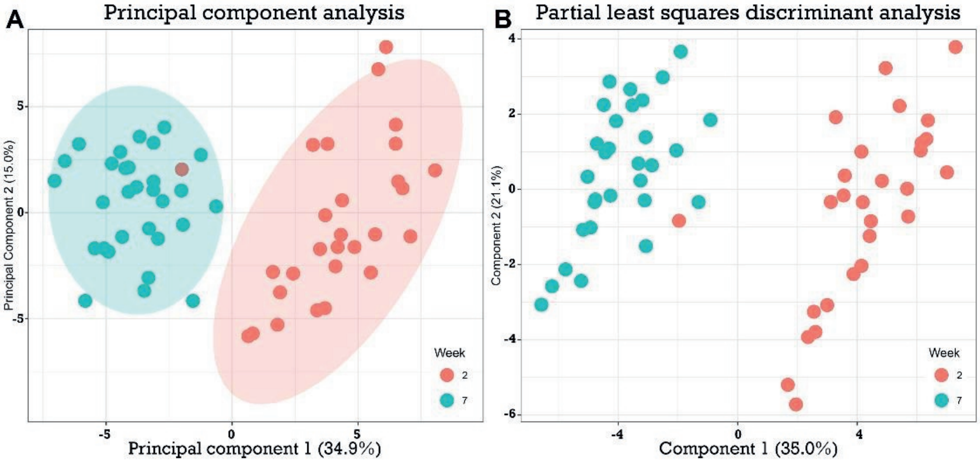


Fig. 2.1. The combined data of milk metabolomics profiles and milk production traits were separated by lactation week in principal component analyses (A), and were further discriminated by lactation week with a discriminant power of $Q^2 = 0.86$ in partial least squares discriminant analysis (B). Numbers in parentheses is the percentage of explainable variation of milk metabolomics profiles and milk production traits due to separation between week 2 and week 7.

2.4.2. Correlation analysis

In the current study, the energy balance of dairy cows was measured, and correlations between energy balance and milk metabolites were analysed. Fig. 2.3 shows Pearson correlations among the top 15 variables with the highest VIP scores for the first principal component in the PLS model to estimate energy balance of dairy cows in lactation week 2 and 7. The complete correlation matrix of 52 milk metabolites and 6 milk production traits with energy balance is shown for lactation week 2 and 7 separately in *Appendix* Fig. 2.2. All top 15 variables in the PLS model were correlated with energy balance in lactation week 2, the correlation coefficient (*r*) ranged from -0.88 to -0.55, or 0.56 to 0.71 (Fig. 2.3-A). The same trend was observed in lactation week 7 (Fig. 2.3-B). In earlier studies, milk metabolites were identified as indicators for milk characteristics and metabolic status of dairy cows. Choline, carnitine, citrate and lactose in milk were correlated with coagulation properties of milk [135]. Increased β-hydroxybutyrate concentration in milk was related to subclinical ketosis of dairy cows [22]. In addition, an observed correlation between an amino acid, lysine, and protein content in milk was proposed to relate to the limiting nature of amino acids for protein production [54]. In our study, five milk production traits, milk yield, fat yield, protein yield, lactose yield and FPCM, were clearly correlated (P-value $< 5.0\times10^{-2}$) with each other in both lactation weeks. The correlation between top 15 variables in milk and energy balance indicated a biological relationship between these milk variables and energy balance, which will be discussed in more detail below.

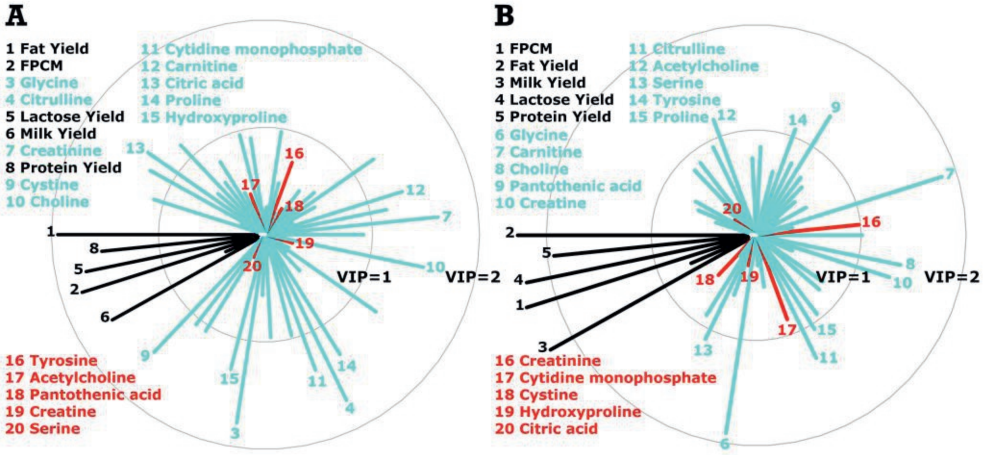


Fig. 2.2. Variable importance in projection (VIP) scores in the first principal component calculated by partial least squares (PLS) to estimate the energy balance of dairy cows in lactation week 2 (A) and week 7 (B). The top 15 metabolites with relatively higher VIP score are shown. The relatively higher VIP score means that variable has higher capability to estimate energy balance in PLS analysis. Black line and text represent milk production traits, blue line and text represent milk metabolites with relatively higher VIP score, and red line and text represent milk metabolites which were in the top 15 in one week, but not in the other week.

Abbreviations: FPCM, fat- and protein-corrected milk production.

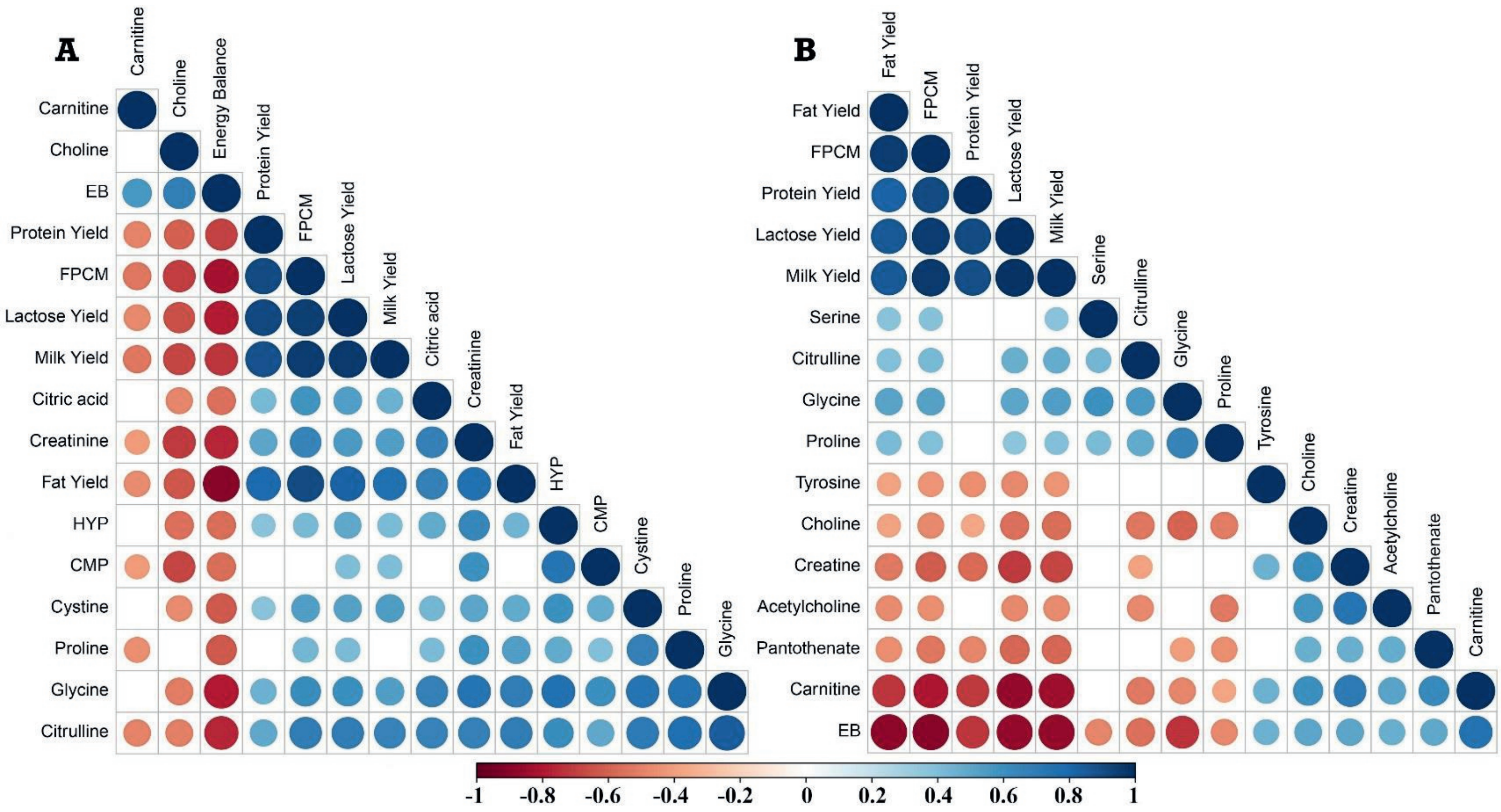


Fig. 2.3. Pearson correlations matrix between 15 variables in milk with relative higher variable importance in projection scores from partial least squares analysis and energy balance (EB) of dairy cows in lactation week 2 (A) and week 7 (B). The size of dots is proportional to the absolute value of correlations, the blue and red colour represent positive and negative direction of correlations, respectively, and blank represents that there was no correlation between two variables (*P*-value > 0.05).

Abbreviations: CMP, cytidine monophosphate; FPCM, fat- and protein-corrected milk; HYP, hydroxyproline.

2.4.3. The effect of dry period length, parity and energy balance on milk metabolites and milk production traits analysis

The effect of DPL, parity and energy balance on milk metabolites and milk production traits of dairy cows in lactation week 2 and 7 was analysed with a mixed model. Briefly, DPL, parity, energy balance and their two-way interactions were included in the model. Significant two-way interactions (P-value < 5.0×10^{-2}) were kept in the model (*Appendix* Fig. 2.3 and 2.4).

Dairy cows with a 0-d dry period had a lower lactose and protein yield than cows with a 30-d dry period in week 2 after calving (Table 2.1), which is in line with earlier studies [86,136]. The lower lactose and protein yield was probably due to the lower milk yield of cows with a 0-d dry period, compared with cows with a 30-d dry period [86]. The effect of DPL on milk yield depended on the energy balance (P-value = 3.1×10^{-3}) of cows, cows with a 30-d had a higher milk yield than cows with a 0-d dry period only when the energy balance was more than -125.0 $kJ/kg^{0.75}\cdot d$. Also, protein yield, lactose yield and FPCM were affected by an interaction between DPL with energy balance of cows (*Appendix* Fig. 2.3). In week 7, dairy cows with a 0-d dry period had a higher concentration of acetyl-choline, pantothenate and tyrosine than cows with a 30-d dry period (Table 2.2). The two-way interaction between DPL and energy balance affected citrulline, glycine, protein yield and FPCM. The effect of DPL had the same direction to milk metabolites or milk production traits on different energy balance states in lactation week 2 and 7 (*Appendix* Fig. 2.3 and 2.4). In earlier studies, the lower milk yield after shortening (28 to 30-d) or omitting (0-d) dry period improved energy balance and metabolic status in dairy cows in early lactation, compared with a conventional dry period of (56 to 60-d) [86,137]. Also, in the current study, cows with a 0-d dry period had a better energy balance than cows with a 30-d dry period (-77.2 vs. -276.8 $kJ/kg^{0.75}$ per day for 0-d vs. 30-d dry period in week 2, P-value = 6.7×10^{-3}; and 112.7 vs. -27.3 $kJ/kg^{0.75}\cdot d$ for 0-d vs. 30-d dry period in week 7, P-value = 2.5×10^{-2}). In lactation week 2, milk yield, protein yield, lactose yield and FPCM were affected by DPL, but this effect was depended on its two-way interaction with energy balance of cows. In week 7, only FPCM and protein yield were affected by DPL, related to its two-way interaction with energy balance of cows.

Table 2.1. The effect of dry period length (DPL), parity and energy balance (EB), on 10 milk metabolites and 5 milk production traits obtained through partial least squares analysis (PLS) in lactation week 2. These 15 variables had a relatively higher VIP score [1] in week 2.

	DPL		**Parity**		**EB**	**Two-way interaction**		
	FC [a]	*P*-value	FC [b]	*P*-value	*P*-value	EB*DPL	EB*Parity	DPL*Parity
Milk metabolites								
Carnitine	1.19	0.80	1.28	0.52	<0.01	NM	NM	NM
Choline	1.41	0.35	1.51	<0.01	<0.01	NM	0.05	NM
Citrate	0.80	0.12	0.93	0.94	<0.01	NM	NM	0.08
Citrulline	0.69	0.92	0.90	0.12	<0.01	NM	NM	NM
Creatinine	0.86	0.94	0.85	0.43	<0.01	NM	NM	0.02
Cystine	0.70	0.68	0.78	0.84	<0.01	NM	NM	NM
CMP	0.62	0.77	0.74	0.14	0.02	NM	0.08	NM
Glycine	0.54	0.33	0.71	0.13	<0.01	0.08	<0.01	0.10
HYP	0.65	0.73	0.90	0.54	<0.01	NM	0.10	NM
Proline	0.94	0.65	0.98	0.88	<0.01	NM	NM	NM
Milk production traits								
Fat (kg)	0.81	0.39	0.85	0.20	<0.01	0.07	NM	0.01
FPCM (kg)	0.85	0.07	0.83	0.55	<0.01	<0.01	NM	0.08
Lactose (kg)	0.87	0.05	0.87	0.14	<0.01	<0.01	NM	NM
Milk yield (kg)	0.88	0.13	0.82	0.84	<0.01	<0.01	NM	NM
Protein (kg)	0.86	<0.01	0.89	0.12	<0.01	<0.01	NM	0.06

[a] Fold change in the metabolite concentration (DPL 0/30).
[b] Fold change in the metabolite concentration (parity 2/3).
[1] Variable importance in projection score in partial least squares analysis.
Abbreviations: CMP, cytidine monophosphate; FPCM, fat- and protein-corrected milk production; HYP, hydroxyproline; NM, not included in model.

Parity had an effect on milk metabolites and milk production traits of dairy cows in both lactation weeks. In week 2, young cows (parity 2) had a higher choline concentration than older cows (parity 3). In week 7, young cows had higher choline, carnitine and pantothenate concentration, but lower glycine concentration, milk yield and lactose yield, compared with older cows. Therefore, parity affected more milk metabolites of dairy cows in week 7 than in week 2. Young cows had a different energy balance than older cows in week 2 (-93.8 *vs*. -249.1 kJ/kg$^{0.75}$·d for young *vs*. older cows, *P*-value = 3.3×10^{-2}). Both young and older cows recovered from NEB in week 7 (67.9 *vs*. 17.5 kJ/kg$^{0.75}$·d for young *vs*. older cows, *P*-value = 4.1×10^{-1}). The different energy balance of young cows in week 2 is explained by the lower milk production (FPCM), compared with older cows (29.6 *vs*. 35.9 kg/d for young *vs*. older cows, *P*-value = 1.7×10^{-2}) at a similar energy intake (120.15 *vs*. 118.04 MJ/d for young *vs*. older cows, *P*-value = 0.67).

Table 2.2. The effect of dry period length (DPL), parity and energy balance (EB), on 10 milk metabolites and 5 milk production traits obtained through partial least squares analysis (PLS) in lactation week 7. These 15 variables had a relatively higher VIP score [1] in week 7.

	DPL		Parity		EB	Two-way interaction		
	FC [a]	*P*-value	FC [b]	*P*-value	*P*-value	EB*DPL	EB*Parity	DPL*Parity
Milk metabolites								
Acetyl-choline	1.73	<0.01	1.07	0.94	0.15	NM	NM	NM
Carnitine	1.21	0.19	1.17	0.03	<0.01	NM	NM	NM
Choline	1.23	0.15	1.27	0.01	<0.01	NM	NM	NM
Citrulline	0.83	0.21	0.97	0.99	<0.01	<0.01	NM	0.03
Creatine	1.17	0.05	1.13	0.05	0.05	NM	NM	NM
Glycine	0.70	0.12	0.76	0.02	<0.01	<0.01	<0.01	NM
Pantothenate	1.32	<0.01	1.18	0.04	0.10	NM	NM	NM
Proline	0.88	0.35	0.85	0.22	0.02	NM	NM	NM
Serine	0.87	0.93	0.87	0.36	0.02	NM	NM	NM
Tyrosine	1.03	0.03	1.05	0.13	<0.01	0.02	<0.01	<0.01
Milk production traits								
Fat (kg)	0.84	0.52	0.92	0.47	<0.01	NM	NM	NM
FPCM (kg)	0.86	0.69	0.88	0.15	<0.01	0.02	NM	NM
Lactose (kg)	0.82	0.86	0.85	0.03	<0.01	0.09	NM	NM
Milk yield (kg)	0.84	0.95	0.83	<0.01	<0.01	NM	NM	NM
Protein (kg)	0.91	0.25	0.89	0.08	<0.01	<0.01	NM	NM

[a] Fold change in the metabolite concentration (DPL 0/30).
[b] Fold change in the metabolite concentration (parity 2/3).
[1] Variable importance in projection score in partial least squares analysis.
Abbreviations: FPCM, fat- and protein-corrected milk production; NM, not included in model.

In each one of lactation week 2 and 7, 10 milk metabolites and 5 milk production traits with the highest VIP scores were related to the energy balance in the PLS models. It could be expected that the severity in NEB is related to the metabolic profile in milk of dairy cows, although little information has been available hitherto on this relationship. In an earlier study, we proposed that the presence of sugar phosphates in milk of cows during severe NEB indicated leakage of these components from mammary epithelial cells into milk due to apoptosis [66].

Model 2, 5 and 8 using both milk metabolites and milk production traits had the highest capacity to estimate the energy balance of dairy cows. In both week 2 and 7, models based on milk metabolites had limited estimating capacity than models based on milk production traits (adjusted-$R^2 = 0.68$ *vs.* 0.78 in week 2; adjusted-$R^2 = 0.65$ *vs.* 0.80 in week 7). In earlier work, milk production traits, including, milk yield, fat, protein and lactose yield, were highly related to the energy output of dairy cows [138,139]. In week

2, dairy cows were suffering from severe NEB, energy balance was -177.0 kJ/kg$^{0.75}$·d, while cows were recovering from NEB in week 7, energy balance was 42.7 kJ/kg$^{0.75}$·d (P-value = 1.5×10^{-5}). The difference in energy balance between lactation weeks indicate that dairy cows were also in different metabolic status in lactation week 2, compared with week 7. A relative higher amount of body reserves can be hypothesized to be mobilized for milk synthesis in week 2 than in week 7, which may explain why the capacity of milk production traits to estimate the energy balance only in week 2 is greater than when both weeks are combined. In earlier studies, milk production traits were used as promising indicators for energy balance at herd level and subclinical ketosis of individual cows [21,140]. Milk fat-to-protein ratio increased coefficients by 19.0% to 52.0% when it was used to estimate herd-level energy balance at herd level [37]. Also in this study, fat yield contributed significantly to the estimation of the energy balance of individual cows (Table 2.3). Fat yield accounted for 65.8% and 63.9% the variation of energy balance in Model 2 and 5, respectively. Increasing fat yield resulted in greater energy output of dairy cows in early lactation [18,27], which accounted for the lower energy balance and the role of fat yield in reduced models to estimate the energy balance where milk production traits were involved. Fat yield in model 8 accounted for 48.8% explained variation for energy balance. Glycine, carnitine and citrate together accounted for more explained variation to estimate energy balance, 34.1%, 8.6% and 8.5%, respectively, than fat yield alone. Compared with Model 2 and 5, the explained variation by fat yield decreased in Model 8. The lower variation by fat yield in Model 8 was in line with the low coefficient to estimate energy balance in Model 9 by milk production traits only. Glycine was an important metabolite to estimate the energy balance in all models with milk metabolites, next to carnitine, choline, tyrosine and citrate. The possible biological relationship between glycine and the energy balance in early lactation is discussed in more detail below. Moreover, reliable and practical tests should be developed to detect these milk metabolites to estimate the energy balance under practical conditions.

2.4.4. Energy balance and biological pathways

In early lactation, dairy cows mobilize their body reserves to meet the energy requirement for milk synthesis and secretion in the mammary gland [141,142]. Therefore, the energy balance of individual cows could be expected to be reflected in metabolic patterns in milk. In previous studies, metabolic changes were observed in early lactation of cows as compared with late lactation [143,144]. Klein *et al.* (2013) observed that 19 amino acids, glucose and a number of carboxylic acids were related to ketosis in dairy cows [143]. Ketosis indicates the NEB in cows, but its relationship with milk metabolites is not fully clear. Klein *et al.* (2013) observed a positive correlation of plasma glycine-to-alanine ratio to milk ketone bodies, acetone ($r = 0.77$) and BHB ($r = 0.64$), suggesting an excessive protein mobilization and a shortage in glucose supply, and it could be the reason for these biological effect on a shortage of vitamin B6 in these animals [143]. Meijer *et al.* (1995) observed large changes in amino acids in muscle and plasma of high yielding dairy cows in early lactation, suggesting that protein was degraded for the supply of amino acids to the udder [144]. It was proposed that glutamine is potentially limiting for protein synthesis. Although these and similar studies have been done carefully, they do not provide a clear biological relationship of energy balance with metabolism of individual cows. In our study, we measured both energy balance and metabolic variables in milk of individual cows.

Table 2.3. Reduced models to estimate the energy balance of dairy cows in lactation week 2 and 7. The reduced models were selected by multivariate linear regression.

Model	No.	Model (Equation)	R^2[1]	ad-R^2[2]
Dataset of dairy cows in week 2				
M[3]	**1**	EB = -357.2 - 1.9*glycine (60.0%) + 0.5*choline (21.6%) + 1.6*carnitine (18.4%)	0.72	0.68
M+P[4]	**2**	EB = 222.0 - 288.2*fat (65.8%) + 0.3*choline (18.0%) - 1.2*glycine (16.1%)	0.85	0.83
P[5]	**3**	EB = 580.7 - 532.4*fat	0.79	0.78
Dataset of dairy cows in week 7				
M	**4**	EB = -204.3 - 3.2*glycine (60.7%) + 1.9*carnitine (29.9%) + 35.2*tyrosine (9.4%)	0.69	0.65
M+P	**5**	EB = 591.3 - 334.2*fat (63.9%) - 2.4*glycine (30.3%) + 28.7*tyrosine (5.9%)	0.89	0.88
P	**6**	EB = 632.2 - 331.4*fat (46.4%) - 14.9*milk yield (37.8%) + 338.9*protein (15.8%)	0.81	0.80
Dataset of dairy cows in both week 2 and week 7				
M	**7**	EB = -178.5 - 2.6*glycine (56.8%) + 2.1*carnitine (43.2%)	0.77	0.76
M+P	**8**	EB = 222.9 - 301.3*fat (48.8%) - 1.7*glycine (34.1%) + 1.1*carnitine (8.6%) + 5.2*citrate (8.5%)	0.88	0.87
P	**9**	EB = 613.4 - 648.0*fat (82.6%) + 653.4*lactose (11.4%) - 21.6*milk yield (6.0%)	0.55	0.53

[1] R^2 was obtained through 10-fold cross-validation.

[2] *ad-R^2* considered the number of independent regressors in a model, and it was obtained through formula, *ad-R^2* = 1 - [(1 - R^2)(n-1)/(n-k-1)], n is the number of sample size, k is the number of independent regressors, excluding the constant.

[3] M: only milk metabolites are used in the model.

[4] M + P: both milk metabolites and milk production traits are used in the model.

[5] P: only milk production traits are used in the model.

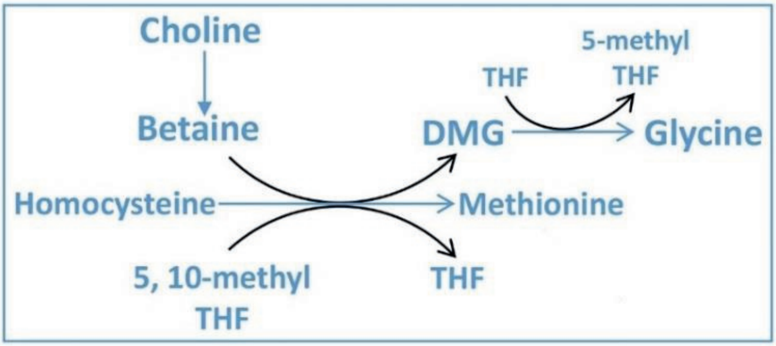


Fig. 2.4. As a major methyl donor from diet, choline transfers its methyl-group to S-adenosyl methionine via betaine with a concomitant formation of glycine in this process. The figure was adapted from Friesen *et al.* (2007) [145].

Abbreviations: THF, tetrahydrofolate; DMG, dimethylglycine.

Of the milk metabolites observed in our study, a large number of metabolites were correlated, either positively or negatively, to the energy balance of the individual cows (Fig. 2.2). Some of these metabolites could be involved in the same metabolic pathway which can be extrapolated from the strong correlations between some components as observed in Fig. 2.3. We therefore created reduced models in order to include metabolites with a limited interdependence (Table 2.3). Of the metabolites in our reduced models, glycine, choline, carnitine, citrate and tyrosine were the ones which were most clearly related to the energy balance. Glycine was found to be the most important metabolite in all models. In previous studies, both plasma and milk glycine concentrations were increased during early lactation of cows, compared with late lactation [143,144]. In the current study, milk glycine was negatively related to energy balance in both lactation week 2 (r = -0.80) and week 7 (r = -0.74) (Fig. 2.3). Besides glycine also choline is related to energy balance (Table 2.3). These two metabolites, glycine and choline, are both important in the one carbon metabolism (Fig. 2.4). In the one carbon metabolism dietary choline is a methyl donor for important biological processes involving the folate cycle, redox balance status and cell renewal [145,146]. Choline is regarded as a limiting nutrient for transition dairy cows [147,148]. In early studies, dairy cows supplied with rumen-protected choline had increased milk production [149], but the decreased incidence of diseases, including fatty liver [150], ketosis and mastitis [151]. Those early studies tested the positive effect of choline on animal healthy and production, however, no energy balance was shown. Our finding that choline deficiency in severe NEB cows could explain the low production and high disease incidence in early lactation of cows.

In our study, choline is not only negatively correlated with glycine, but also positively correlated with carnitine (Fig. 2.3). The question is how these three small metabolites, glycine, choline and carnitine are mechanistically related to energy balance. Cows with severe negative energy balance have high concentrations of glycine in milk, and these cows have low levels of choline and carnitine in milk (*Appendix* Fig. 2.3). In cows with high energy balance the opposite is observed for these three metabolites, low levels of glycine and high levels of choline and carnitine. Carnitine is a quaternary ammonium salt as is choline. Carnitine is involved in fatty acid metabolism (Fig. 2.5) and is not directly involved in one carbon metabolism. However, as carnitine is synthesized in the liver from lysine and methionine a possibly relationship to one carbon metabolism could exist. On the other hand, it has been proposed that a major source of carnitine originates from protein lysine N-trimethylation on release from proteins by protein hydrolysis [152]. The low levels of carnitine in negative energy balance cows could therefore be the result of low carnitine synthesis in the liver or from low levels of protein hydrolysis. However, low levels of carnitine could be also the result of high levels of acyl-carnitines. Formed in lipid and amino acid oxidation processes, acyl-carnitines are intermediates in the breakdown processes of lipids and amino acids. Giesbertz *et al.* (2015) quantitatively measured the concentration of 56 acyl-carnitines in mice with a metabolic syndrome [153]. Unfortunately, we did not measure acyl-carnitines in our study and no relationship between acetyl-carnitine, carnitine or energy balance was found. It can be expected that more lipid metabolism occurs in cows with NEB, which consumed carnitine and generated various acyl-carnitines. Future studies have to show if the low levels of carnitine in negative energy balance cows are related to low levels of biosynthesis of carnitine or high levels of acyl-carnitines. From our findings that in severe negative energy balance glycine levels are relatively high in concentration and methylated metabolites as choline, and carnitine are low in concentration, we conclude that severe NEB cows have a need for methyl donor compounds. This observation mimics observations made in cancer cells, where there is a high demand for methyl donors [154,155]. Labuschagne *et al.* (2014) observed that in cancer cells there is a high demand for methyl donors and that these methyl groups are used in nucleic acid synthesis [154]. Interestingly, glucose was observed in our study to be related to energy balance (low levels of glucose in NEB cows) and we propose that glucose is used additionally as methyl donor via the glucose-serine pathway, rather than only for lactose production

[155]. Glucose and glucogenic amino acids were regarded as the precursor of milk production via propionate in early lactation of cows [87,156].

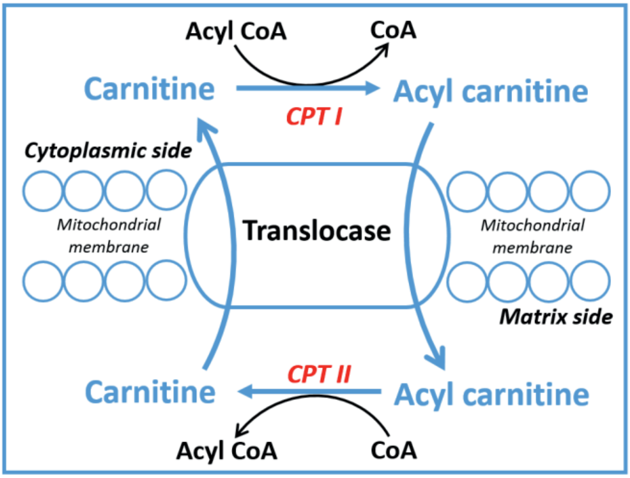


Fig. 2.5. Fatty acid transportation mechanism in the cell. The inner mitochondrial membrane is impermeable to fatty acids and a specialized carnitine carrier system operates to transport activated fatty acids from cytosol to mitochondria. Carnitine is converted to acyl carnitine for fatty acid transportation. The figure was adapted from Nelson *et al.* (2008) [157].

Abbreviations: CoA, coenzyme A; CPT I, carnitine palmitoyltransferase I; CPT II, carnitine palmitoyltransferase II.

Finally, we also observed relationships between amino acids and negative energy balance. These amino acids are highly correlated and in our reduced model tyrosine was selected. In the current study, cows with a NEB had low levels of tyrosine in milk. As the energy balance was negatively correlated with protein yield in both lactation week 2 and 7, we propose that the lower levels of amino acids in milk are related to higher levels of protein synthesis.

2.5. Conclusions

In the current study, 52 milk metabolites of dairy cows during a status of negative energy balance were detected through LC-QQQ-MS. The energy balance had a high correlation with specific milk metabolites and milk production traits in both lactation week 2 and 7. Nine reduced models were composed to estimate the energy balance of dairy cows in lactation week 2 and 7, with a range from 53% to 88% predicting power. Both milk metabolites and milk production traits had an important role in these models, in particular glycine, choline, carnitine and fat yield. The strong relationship of these metabolites with energy balance of dairy cows in early lactation could be related to their roles in cell renewal.

CHAPTER 3

Mammary Gland Metabolism of Dairy Cows in Early Lactation

Wei Xu[1,2], Ariette van Knegsel[1], Edoardo Saccenti[3], Renny van Hoeij[1], Bas Kemp[1] & Jacques Vervoort[2]

1. Adaptation Physiology Group, Department of Animal Sciences, Wageningen University & Research, the Netherlands

2. Laboratory of Biochemistry, Wageningen University & Research, the Netherlands

3. Laboratory of Systems and Synthetic Biology, Wageningen University & Research, the Netherlands

Submitted

3.1. Abstract

Milk production of dairy cows increases progressively in the weeks after calving. Consequently, dairy cows experience a negative energy balance (NEB) in early lactation because feed intake is too low to meet energy requirements for body maintenance and milk production. A NEB is expected to affect the metabolism in the mammary gland, which could be reflected by the metabolic profile of milk. In this study our aims were to measure the metabolic profile of milk of dairy cows in lactation week 2 using liquid chromatography mass spectrometry (LC-MS) and nuclear magnetic resonance (NMR) and to investigate the effect of energy balance on the metabolism in the mammary gland. Nuclear magnetic resonance and LC-MS techniques are complimentary techniques that enabled a comprehensive overview of milk metabolites in our study. Energy balance and milk samples were obtained from 87 dairy cows. A total of 67 milk metabolites were detected. Of these 67 milk metabolites, 15 were positively correlated to energy balance and 26 were negatively correlated to energy balance. Cows in more serious NEB produced more milk with increased milk fat yield and higher concentrations of citrate, cis-aconitate, creatinine, glycine, phosphocreatine, galactose-1-phosphate, glucose-1-phosphate, UDP-N-acetyl-galactosamine, UDP-N-acetyl-glucosamine and phosphocholine, but lower concentrations of choline, ethanolamine, fucose, N-acetyl-neuraminic acid, N-acetyl-glucosamine and N-acetyl-galactosamine. We conclude, based on these observations, that the metabolic processes in the mammary gland during NEB are related to the leakage of cellular content, the synthesis of nucleic acids, the synthesis of cell membrane phospholipids, protein glycosylation, an increase in one-carbon metabolic processes as well as an increase in lipid-triglyceride anabolism. Overall, we concluded that both apoptosis combined with cell renewal are paramount in the mammary gland in cows in NEB, which might be related to the high milk yield of the cows during NEB.

3.2. Introduction

In early lactation of dairy cows, elevated energy requirements for milk production combined with a relatively low dry matter intake (DMI) result in an energy deficit or negative energy balance (NEB) [27,28]. A severe NEB is related to an increased risk of metabolic disorders and diseases, such as fatty liver and ketosis [158,159]. In previous studies it was shown that body reserves were mobilized to meet the nutritional demand of the mammary gland for milk production with concurrent changes in metabolic hormones and plasma metabolites [105,160]. In the mammary gland, the basic metabolic patterns related to the synthesis of milk fat, protein, and lactose are known [161,162]. However, the modifications of metabolic pathways, especially in the relation to negative energy balance, have not been described yet.

Metabolomics studies aim to detect and quantify small molecules from bio-fluids through several metabolomics techniques, such as mass spectrometry (MS) and nuclear magnetic resonance (NMR). Integrating the results of different metabolomics techniques allows new insights into the metabolic profiles [118-120]. Liquid chromatography mass spectrometry (LC-MS) is a powerful technique with high sensitivity and selectivity [163]. High resolution NMR is a very stable technique with better reproducibility than LC-MS but NMR suffers, relative to LC-MS, from limited sensitivity [164]. In recent years, integrated analyses that combines results from LC-MS and NMR have been applied to detect and quantify a wide range of metabolites in bio-fluids, such as urine, plasma and milk [57,143,165,166]. The integration of data from different techniques supports cross-assigning signals from the techniques on the same samples [166]. In dairy cows, the integration of MS and NMR data identified biomarkers of heat stress in plasma [57], and investigated the correlation of plasma and milk metabolites [143]. In the past decade, new developments with hydrophilic columns make analysis of polar metabolites possible using LC-MS. We used a recently introduced pentafluorophenylpropyl column to separate polar metabolites with subsequent identification and quantification using a triple-quadrupole-MS. These results combined with high resolution NMR measurements of the same samples has the potency to detect and quantify more metabolites in milk than in any study before.

With the combined NMR and LC-MS datasets we were able to better understand biological pathways affected by NEB and associated alterations in metabolic status of dairy cows. In our study, 87 dairy cows were studied. In lactation week 2, milk yield, milk composition (fat, protein, and lactose), DMI, and energy balance of individual cows were recorded, as well as milk samples were collected. From our data we could obtain a detailed metabolic pattern occurring in cows with severe negative energy balance.

3.3. Materials and Methods

3.3.1. Animals and experimental design

The experimental protocol for the study was approved by the Institutional Animal Care and Use Committee of Wageningen University and was conducted at Dairy Campus research farm (WUR Livestock Research, Lelystad, the Netherlands). Cows and samples in the current study originated from an experiment studying effects of dry period length and dietary energy level on milk yield and energy balance including 123 cows, which was described previously [126]. Earlier, we reported the relation of energy balance and metabolites detected in milk samples of 31 dairy cows in lactation week 2 and 7 through LC-MS measurement (Chapter 2), those 31 cows were not included in the current study. Samples and data of 91 dairy cows were available for the current study. Of these 91 cows, 4 cows were omitted from the analysis because 2 cows suffered from clinical mastitis, 1 cow suffered from metritis, and 1cow had locomotion problems in sampling week. Resulting in total 87 high-yielding Holstein-Friesian dairy cows averaging 663.5±67.6 kg of body weight (in lactation week 2 after calving) participated. Dairy cows were blocked for parity, expected calving period and expected milk yield. Within blocks, cows were randomly assigned to one of two dry period length (DPL, 0 day, 2/3 of the cows; or 30 days: 1/3 of the cows) before calving. Prepartum, cows with a 0 day DPL received a lactation ration based on grass silage and corn silage (6.4 MJ net energy for lactation (NE)/kg dry matter (DM)). Cows with a 30 days DPL received a dry cow ration based on grass silage, corn silage and wheat straw (5.4 MJ NE/kg DM). Postpartum, all cows received the same basal lactation ration as provided to lactating cows prepartum plus additional concentrates. Postpartum, concentrate supply increased stepwise with 0.3 kg/d till 8.5 kg/d on 28 DIM. Body weight, milk yield and feed intake were recorded daily. During lactation, cows were milked twice daily at ~0600 hours and ~1800 hours.

3.3.2. *Milk samples*

Milk samples for fat, protein, and lactose percentage (ISO 9622, Qlip, Zutphen, The Netherlands) were collected four times per week (Tuesday afternoon, Wednesday morning, Wednesday afternoon, and Thursday morning), and then stored at -20 °C until analysis. Milk samples were analysed as a pooled sample per cow per week and used to calculate average fat, protein and lactose yield in this week. Milk samples for metabolomics analysis were collected at Wednesday morning in lactation week 2. Milk production traits were averaged per week. Four milk samples were omitted from the analysis because two cows suffered from clinical mastitis, one cow suffered from metritis, and one cow had locomotion problems in sampling week. Fat- and protein-corrected milk was calculated as:

$$FPCM\ (kg) = [0.337 + 0.116 \times fat\ (\%) + 0.06 \times protein\ (\%)] \times milk\ yield\ (kg)\ [167].$$

3.3.3. *Energy intake and energy balance*

Roughage and concentrate were supplied separately, daily intakes were recorded per individual cow using roughage intake control troughs (Insentec, Marknesse, the Netherlands). Energy balance was calculated per week according to the Dutch net energy evaluation (VEM) system, as the difference between net energy intake and the estimated net energy requirements for maintenance, and milk yield (1,000 VEM = 6.9 MJ of NE) [33,168].

3.3.4. *NMR measurement and data pre-processing*

Sample preparation and NMR measurement were performed as earlier described [66,169]. Briefly, milk samples were first thawed to room temperature. The fat layer of milk was removed by addition of deuterated chloroform and centrifugation (12,000 rpm, 15 min, Centrifuge 5424, Eppendorf). Subsequently 175 µL of milk serum was mixed with 175 µL of phosphate buffer (pH = 7.0) and these samples were filtered to remove protein using a Pall 0.5 mL 10-kDa cut-off spin filter with centrifugation at 12,000 rpm for 15 min. Samples were measured with a 3 mm NMR tube (Bruker matching system) using a Bruker NMR spectrometer Avance III with a 600 MHz/54 mm UltraShielded Plus magnet equipped with a CryoPlatform cryogenic cooling system, a BCU-05 cooling unit and an ATM automatic tuning and matching unit. Measurements were done at 300 K. 1D nuclear Overhauser enhancement spectroscopy (NOESY) spectra were obtained.

Baseline corrections, alignment to the resonance of alanine (1.484 ppm) and calibration to internal maleic acid were done for all spectra. Assignment of metabolites resonances was performed using published literature, the Human Metabolome Database version 2.0 online library (http://hmdb.ca/) as well as internal standards.

3.3.5. LC-MS measurements

For quantification of metabolites, a targeted, standardized and quality controlled metabolic phenotyping was performed based on LC-QQQ-MS analysis. The sample as prepared for NMR was also used for analysis with the triple quadrupole mass spectrometer (Shimadzu LC-QQQ-MS; LC-MS-8040) using the PFPP method as described earlier [127,128]. The sample injection volume used was 1 µL, and a single analysis took 25 minutes. From LC-QQQ-MS spectra, metabolites were regarded as reliably identified, when more than 60% observations in all samples showed a reliable value for a certain metabolite.

3.3.6. Integrated analysis and software

The NMR datasets were aligned, the water region was removed before calculating the intensity of the NMR spectra into 0.01 ppm bins. The NMR data (intensity of the bins) were subsequently correlated to energy balance variables. Bins which correlated well to energy balance were selected and the corresponding NMR resonances (peaks) were specifically integrated by carefully selecting peaks that did not show overlap in the NMR spectra. The correlation matrix of NMR and LC-MS datasets was subsequently analysed and edited. In case a metabolite could be identified from both LC-MS and NMR, the intensity of this metabolite was quantified based on its NMR spectrum. Integrated analysis was done as described earlier [166,170]. Pearson correlation coefficient (*r*) and corresponding P value were obtained by function “cor.test ()” in R environment (version 3.3.3). All figures were plotted in R environment (version 3.3.3) with package “ggplot2”.

3.4. Results and Discussion

3.4.1. Measurement by LC-MS and NMR and integrated analysis

Milk samples of 87 dairy cows were measured using both LC-MS and NMR. In LC-MS spectra of milk, 97 metabolites were initially targeted. Overall 39 milk metabolites were detected in LC-MS (*Appendix* Table 3.1). In the NMR spectra of milk, lactose dominated the region around 3.52 to 3.95 ppm, which masked the signal of other metabolites in this region (Fig. 3.1). Therefore, glycine is impossible to be detected by NMR of milk samples at its chemical shift is at 3.54 ppm [171]. Nevertheless, many NMR resonances could be clearly observed resulting in 35 milk metabolites to be detected using NMR (*Appendix* Table 3.1).

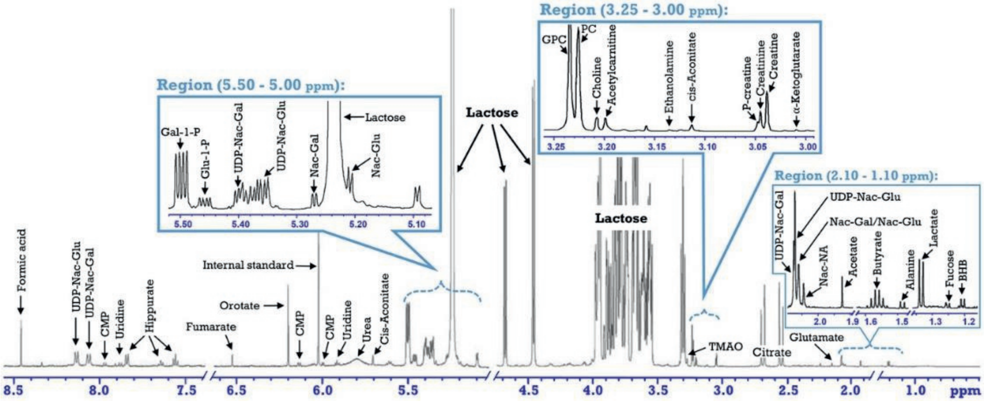
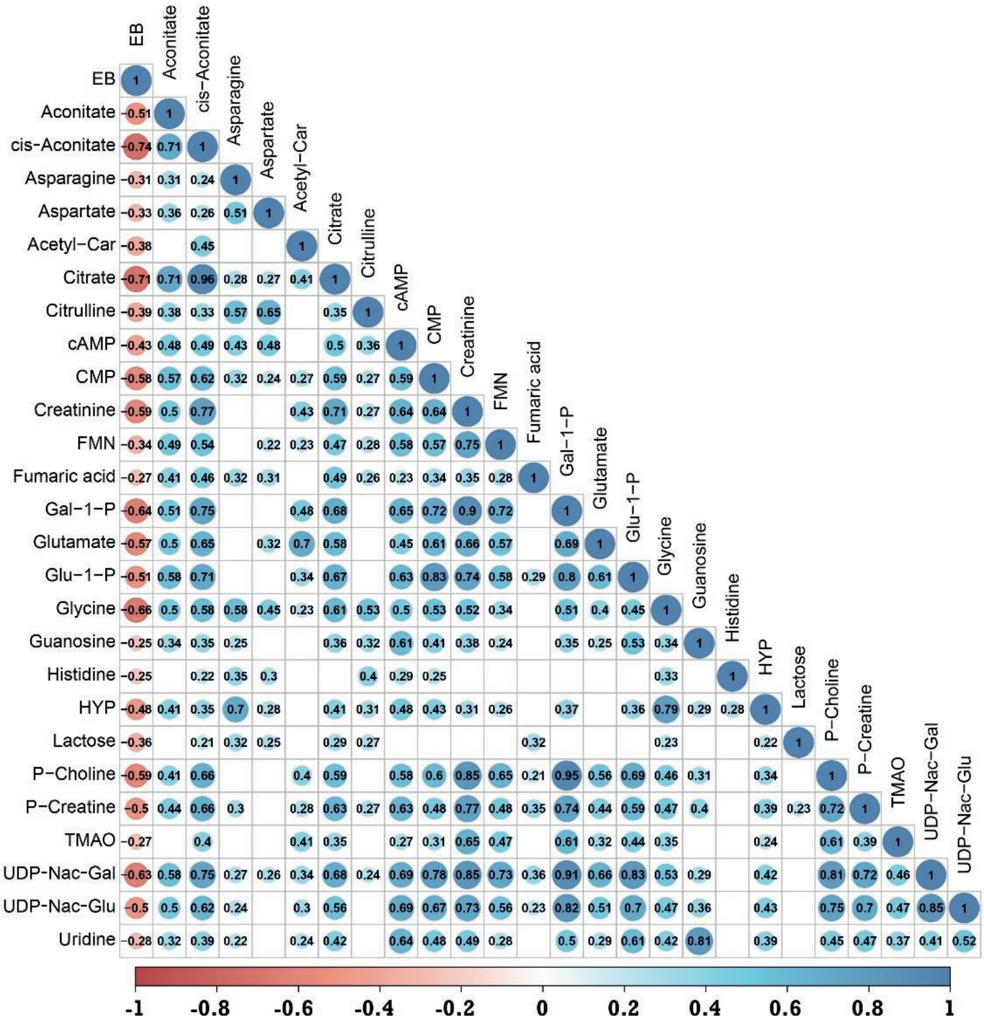


Fig. 3.1. Proton nuclear magnetic resonance spectrum of a milk sample.

Abbreviations: BHB, β-hydroxybutyrate; CMP, cytidine monophosphate; Gal-1-P, galactose-1-phosphate; Glu-1-P, glucose-1-phosphate; GPC, glycerophosphocholine; Nac-Gal, N-acetyl-glucosamine; Nac-Glu, N-acetyl-glucosamine; Nac-Na, N-acetyl-neuraminic acid; PC, phosphocholine; P-creatine, phosphocreatine; phosphocholine; TMAO, trimethylamine N-oxide; UDP-Nac-Gal, uridine diphosphate-N-acetyl-galactosamine; UDP-Nac-Glu, uridine diphosphate-N-acetyl-glucosamine.

A number of milk metabolites have been reported before using LC-MS or NMR [54,64,125]. We integrated LC-MS and NMR datasets. Milk metabolites detected both by NMR and LC-MS had a high correlation between the two measurements, *i.e.* acetyl-carnitine (r = 0.90), choline (r = 0.92), cytidine monophosphate (CMP, r = 0.90), glutamate (r = 0.94), α-ketoglutarate (r = 0.90), uridine (r = 0.93), and valine (r = 0.90). The consistency between two measurements methods indicated that data obtained were reliable.

Through the integration of LC-MS and NMR, 67 metabolites were detected from milk samples of 87 dairy cows in lactation week 2. Of these 67 milk metabolites, 26 were negatively correlated to energy balance (Fig. 3.2), and 15 were positively correlated to energy balance (Fig. 3.3).



	EB	Aconitate	cis−Aconitate	Asparagine	Aspartate	Acetyl−Car	Citrate	Citrulline	cAMP	CMP	Creatinine	FMN	Fumaric acid	Gal−1−P	Glutamate	Glu−1−P	Glycine	Guanosine	Histidine	HYP	Lactose	P−Choline	P−Creatine	TMAO	UDP−Nac−Gal	UDP−Nac−Glu
EB	1																									
Aconitate	-0.51	1																								
cis−Aconitate	-0.74	0.71	1																							
Asparagine	-0.31	0.31	0.24	1																						
Aspartate	-0.33	0.36	0.26	0.51	1																					
Acetyl−Car	-0.38		0.45			1																				
Citrate	-0.71	0.71	0.96	0.28	0.27	0.41	1																			
Citrulline	-0.39	0.38	0.33	0.57	0.65		0.35	1																		
cAMP	-0.43	0.48	0.49	0.43	0.48		0.5	0.36	1																	
CMP	-0.58	0.57	0.62	0.32	0.24	0.27	0.59	0.27	0.59	1																
Creatinine	-0.59	0.5	0.77			0.43	0.71	0.27	0.64	0.64	1															
FMN	-0.34	0.49	0.54		0.22	0.23	0.47	0.28	0.58	0.57	0.75	1														
Fumaric acid	-0.27	0.41	0.46	0.32	0.31		0.49	0.26	0.23	0.34	0.35	0.28	1													
Gal−1−P	-0.64	0.51	0.75			0.48	0.68		0.65	0.72	0.9	0.72		1												
Glutamate	-0.57	0.5	0.65		0.32	0.7	0.58		0.45	0.61	0.66	0.57		0.69	1											
Glu−1−P	-0.51	0.58	0.71			0.34	0.67		0.63	0.83	0.74	0.58	0.29	0.8	0.61	1										
Glycine	-0.66	0.5	0.58	0.58	0.45	0.23	0.61	0.53	0.5	0.53	0.52	0.34		0.51	0.4	0.45	1									
Guanosine	-0.25	0.34	0.35	0.25			0.36	0.32	0.61	0.41	0.38	0.24		0.35	0.25	0.53	0.34	1								
Histidine	-0.25		0.22	0.35	0.3			0.4	0.29	0.25							0.33		1							
HYP	-0.48	0.41	0.35	0.7	0.28		0.41	0.31	0.48	0.43	0.31	0.26		0.37		0.36	0.79	0.29	0.28	1						
Lactose	-0.36		0.21	0.32	0.25		0.29	0.27					0.32				0.23			0.22	1					
P−Choline	-0.59	0.41	0.66			0.4	0.59		0.58	0.6	0.85	0.65	0.21	0.95	0.56	0.69	0.46	0.31		0.34		1				
P−Creatine	-0.5	0.44	0.66	0.3		0.28	0.63	0.27	0.63	0.48	0.77	0.48	0.35	0.74	0.44	0.59	0.47	0.4		0.39	0.23	0.72	1			
TMAO	-0.27		0.4			0.41	0.35		0.27	0.31	0.65	0.47		0.61	0.32	0.44	0.35			0.24		0.61	0.39	1		
UDP−Nac−Gal	-0.63	0.58	0.75	0.27	0.26	0.34	0.68	0.24	0.69	0.78	0.85	0.73	0.36	0.91	0.66	0.83	0.53	0.29		0.42		0.81	0.72	0.46	1	
UDP−Nac−Glu	-0.5	0.5	0.62	0.24		0.3	0.56		0.69	0.67	0.73	0.56	0.23	0.82	0.51	0.7	0.47	0.36		0.43		0.75	0.7	0.47	0.85	1
Uridine	-0.28	0.32	0.39	0.22		0.24	0.42		0.64	0.48	0.49	0.28		0.5	0.29	0.61	0.42	0.81		0.39		0.45	0.47	0.37	0.41	0.52

Fig. 3.2. Pearson correlations matrix among milk production traits, metabolites detected either in NMR, or in LC-MS with energy balance (EB) of dairy cows in lactation week 2. Total 26 milk metabolites were negatively correlated to energy balance.

Abbreviations: Acetyl-Car, acetyl-carnitine; cAMP, 3',5'-cyclic adenosine monophosphate; CMP, cytidine monophosphate; FMN, flavin mononucleotide; Gal-1-P, galactose-1-phosphate; Glu-1-P, glucose-1-phosphate; HYP, hydroxyproline; P-Choline, phosphocholine; P-Creatine, phosphocreatine; TMAO, trimethylamine N-oxide; UDP-Nac-Gal, uridine diphosphate-N-acetyl-galactosamine; UDP-Nac-Glu, uridine diphosphate-N-acetyl-glucosamine.

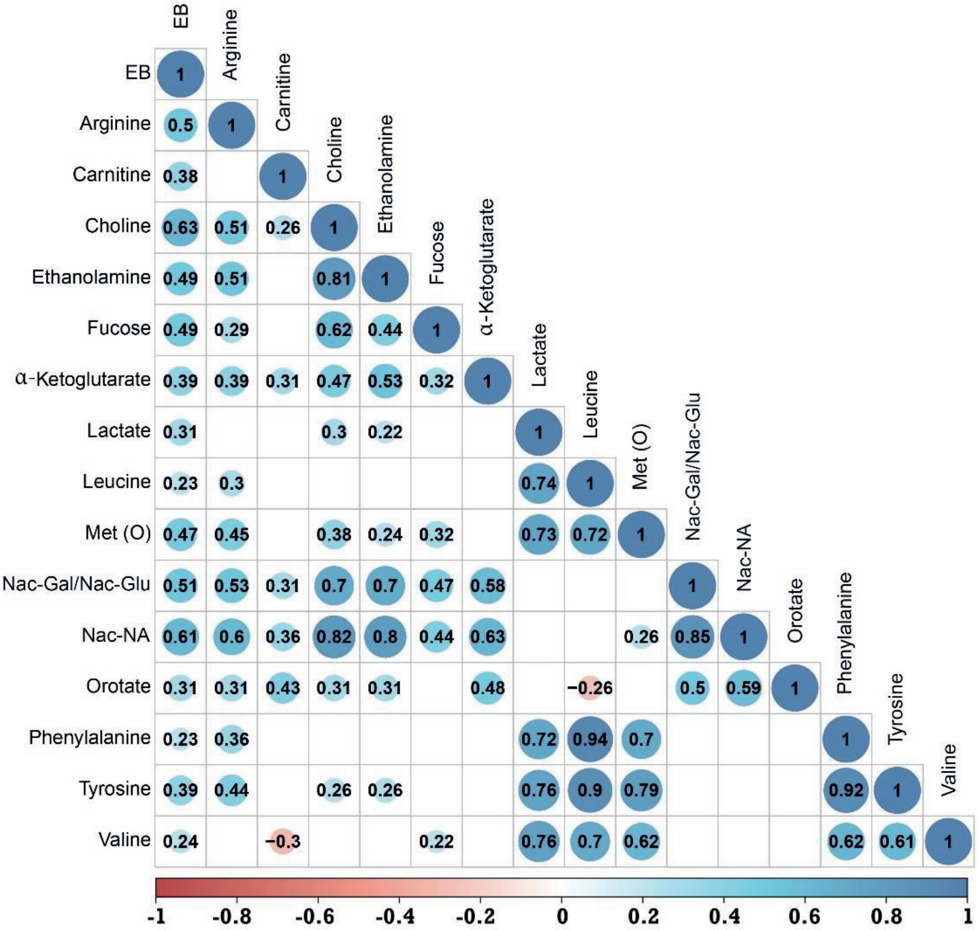


Fig 3.3. Pearson correlations matrix among milk production traits, metabolites detected either in NMR, or in LC-MS with energy balance (EB) of dairy cows in lactation week 2. Total 15 milk metabolites were positively correlated to energy balance.

Abbreviations: Met (O), methionine sulfoxide; Nac-Gal, N-acetyl-galactosamine; Nac-Glu, N-acetyl-glucosamine; Nac-NA, N-acetyl-neuraminic acid.

3.4.2. Cell apoptosis and proliferation

Glucose-1-phosphate (Glu-1-P) and galactose-1-phosphate (Gal-1-P), two intermediates in lactose synthesis, were negatively correlated with energy balance, *r* was -0.51 and *r* was -0.64, respectively (Fig. 3.2). The final step in lactose synthesis is a process confined to the Golgi apparatus [172]. Lactose concentration in milk is always very constant because the water uptake into the secretory vessels within the Golgi apparatus is driven by osmotic force related to the lactose concentration within the vessels [173]. The presence of high concentrations of intermediates in the lactose biosynthesis process in milk indicates that in the mammary gland some cells are leaking cellular content into the milk pool due to apoptosis [66]. Apoptosis in the mammary gland could be related or even caused by low plasma IGF-1 concentrations in dairy cows in NEB [174]. In mammals, IGF-1 is a cell survival factor and an anti-apoptotic factor [175,176]. In dairy cows, the apoptotic index in the mammary gland has been reported be up to 4-fold greater in early lactation than in later lactation [177]. After calving, the substantially increased milk production could result in excessive metabolic stress resulting in apoptosis of epithelial cells in the mammary gland [178,179]. The elevated apoptosis could also be due to discarding non-functional or senescent cells or to removal of a surplus of newly synthesized cells [179]. Besides metabolites related to lactose synthesis, a series of intermediates used for nucleic acids synthesis, present in high concentrations, were observed to have negative correlations with energy balance, such as, uridine (r = -0.28), 3',5'-cyclic adenosine monophosphate (r = -0.43), CMP (r = -0.58), and glycine (r = -0.66) as shown in Fig. 3.2. Cell renewal requires extensive DNA and RNA synthesis and a negative correlation between intermediates in nucleic acids synthesis and energy balance could indicate accelerated cell proliferation in dairy cows in NEB. In the mammary gland of dairy cows, the total DNA content increases by 65% around 10 days pre- and post-partum [180]. The process of cell proliferation increases the number of mammary epithelial cells critical for the increase in milk production in early lactation [161] which is a main contributor to the NEB status of dairy cows during this phase of lactation.

In an eukaryotic membrane, phosphatidylcholine (PtC) and phosphatidylethanolamine (PtE) account for more than 50% of the total phospholipids [181]. The pathway of PtC synthesis from choline and PtE synthesis from ethanolamine using cytidine coenzymes is referred to as the Kennedy pathway [182]. In our study, energy balance was positively correlated with choline (r = 0.63) and ethanolamine (r = 0.49) as shown in Fig. 3.3. In contrast phosphocholine (PC) was observed to be negatively correlated with energy balance (r = -0.59, Fig. 3.2). The rate limiting step in PtC synthesis is the formation of CDP-choline from PC by CTP-phosphocholine cytidylyltransferase (PCT) [183]. The low amounts of choline and the high amounts of PC in dairy cows in NEB indicates that PtC biosynthesis is increased. Remarkably, we observed that PC and choline concentrations were strongly correlated to several metabolites involved in the glycosylation of proteins, N-acetyl-galactosamine (Nac-Gal), N-acetyl-glucosamine

(Nac-Glu), UDP-N-acetyl-galactosamine (UDP-Nac-Gal), UDP-N-acetyl-glucosamine (UDP-Nac-Glu), and N-acetyl-neuraminic acid (Fig 3.4). In phospholipid membrane synthesis proteins present in the membrane need to be glycosylated in order to obtain cellular stability, for signal transduction processes and for viral or microbial defence. In the current study, high amounts of UDP-Nac-Gal, and UDP-Nac-Glu, and low amounts of fucose, N-acetyl-neuraminic acid, Nac-Gal and Nac-Glu were detected in milk of cows in NEB. Milk proteins are heavily glycosylated with fucose, N-acetyl-neuraminic acid, Nac-Gal and Nac-Glu residues as major substituents [184]. UDP-N-acetyl-galactosamine and UDP-Nac-Glu are activated substrates used for protein glycosylation. Apparently, the increased amounts of UDP-Nac-Gal and UDP-Nac-Glu in dairy cows in NEB indicates that there is a high demand for protein glycosylation. The observation that these two UDP-derivatives used for glycosylation are strongly correlated with choline and PC suggests that not only synthesis of membrane phospholipids, but also glycosylation of membrane proteins are of high importance for cows in NEB. Glycosylation of milk serum proteins could be related to glycosylation of membrane proteins as the protein concentration in milk was observed to be strongly correlated with Nac-Gal, Nac-Glu, fucose, N-acetyl-neuraminic acid and choline. Possibly the glycosylation of membrane proteins is the driving force for glycosylation of cytosolic milk proteins. It has been observed that glycosylation of milk proteins is varying depending on lactation week postpartum [185,186], but differences in glycosylation patterns of cytosolic and membrane proteins of individual cows related to energy status has not been studied in detail.

The synthesis of PtC from PC is the rate limiting step in PtC synthesis, possibly because cytidine triphosphate (CTP) is a rate-limiting metabolite in this process (Fig. 3.4). Cytidine triphosphate is not only used for the formation of PtC and PtE in the Kennedy pathway, but is also used to synthesize nucleic acids. In addition, CTP is synthesized from uridine diphosphate (UTP) which is an intermediate in the synthesis of lactose. Dairy cows in NEB had a greater overall lactose yield ($r = -0.59$), which indicates that large amounts of UTP were used to synthesize lactose, competition with the formation of CTP from UTP. Finally, CTP and UTP can be used for DNA and RNA synthesis during cell proliferation [187] creating a huge demand for both UTP and CTP for cows in NEB.

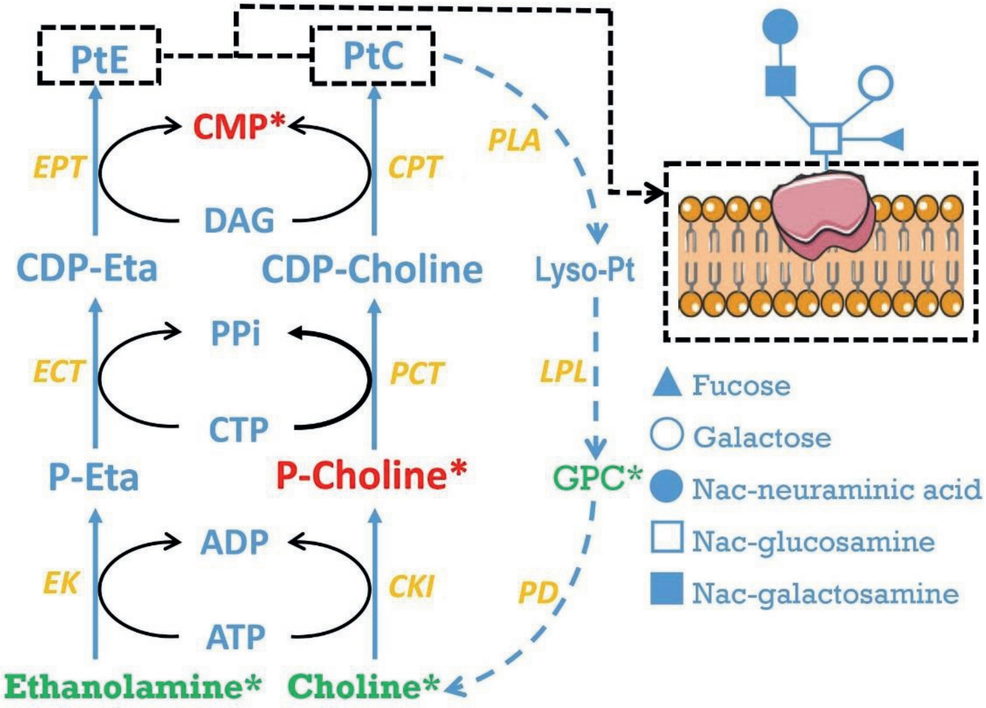


Fig. 3.4. The composition of head groups of a phospholipid bilayer of a cellular membrane. Choline and ethanolamine are two main substrates for the synthesis of phosphatidyl-choline (PtC) and phosphatidyl-ethanolamine (PtE), respectively. Phosphatidyl-choline can be degraded via GPC to choline. The surface proteins of a cellular membrane are glycosylated.

Abbreviations: ADP, adenosine diphosphate; ATP, adenosine triphosphate; CDP, cytidine diphosphate; CKI, choline kinase; CMP, cytidine monophosphate; CPT, 1,2-diacylglycerol cholinephosphotransferase; CTP, cytidine triphosphate; DAG, diglyceride; ECT phosphoethanolamine cytidylyltransferase; EK, ethanolamine kinase; EPT, ethanolaminephosphotransferase; GPC, glycerophosphocholine; LPL, lysophospholipase; P-Choline, phosphocholine; PCT, phosphocholine cytidylyltransferase; PD, glycerophosphocholine phosphodiesterase; P-Eta, phosphoethanolamine; PLA, phospholipase A2; PPi, pyrophosphate.

3.4.3. Fatty acids metabolism

For the synthesis of phospholipids diglycerides are needed (Fig. 3.4). A diglyceride molecule (DAG) contains 2 molecules of fatty acids and 1 molecule of glycerol. In cows in NEB a competition between DAG used for the synthesis of milk triglycerides and DAG used for the synthesis of phospholipids can occur, as triglycerides amounts in milk of dairy cows in NEB are increased. Glycerophosphocholine (GPC) is a product from the breakdown of PtC (Fig. 3.4). A low GPC-to-PC (GPC/PC) ratio was observed for cancer cells in humans [188,189] A low GPC/PC ratio was used to indicate a risk of ketosis in dairy cows [64]. In our study the GPC/PC ratio is positively correlated to energy balance ($r = 0.41$) indicating low levels of GPC/PC in NEB. However this ratio is more the consequence of the high correlation of PC to the energy balance (r = -0.59, indicating high amounts of PC in NEB) than the correlation of GPC to energy balance where a very weak relationship was observed (r = 0.18, P = 0.09). The low amounts of GPC, as well as the low amounts of fucose, N-acetyl-neuramic acid, Nac-Gal, Nac-Glu are in our view related to the reprocessing of cellular components through the lysosome enabling cell proliferation. Therefore, the correlation of GPC/PC to energy balance is related to the process of cell membrane synthesis during cell proliferation in the mammary gland leading to high concentrations of PC.

Citrate detected from NMR spectra was negatively correlated with energy balance (r = -0.71, Fig. 3.2). Citrate is an important metabolite involved in the energy metabolism of cell. In mitochondria, citrate is an intermediate in the tricarboxylic acid cycle where citrate can be isomerized into cis-aconitate. Cis-aconitate is also very well correlated with energy balance (r = -0.74). These observations indicate that citrate is also used extramitochondrial to form acetyl-CoA, and acetyl-CoA can be used for fatty acids synthesis [190]. Mammary epithelial cells are impermeable to citrate in both directions, indicating that citrate is produced intracellular [191]. In the current study, energy balance was negatively correlated with milk fat yield (r = -0.78) and milk fat yield was positively correlated to citrate levels (r = -0.53), indicating that citrate is primarily used for milk fat synthesis in cows in NEB. Citrate levels could be an indicator of energy status [65,192]. Citrate in milk is easily detected. The concentration of milk citrate has a wide variation throughout lactation [193]. Dairy cows have a greater concentration of milk citrate in early lactation than in mid lactation [65], which could be explained by the improved energy balance due to lower milk and milk fat synthesis in mid lactation.

3.5. Conclusions

In this study, 67 metabolites were detected and reliably quantified from milk serum of dairy cows in lactation week 2, using NMR and LC-MS through an integrated analysis. A large number of metabolites (26) were negatively and 15 metabolites were positively related to the energy balance of cows. Based on these data we concluded that apoptosis and cellular proliferation occurs in cows with NEB with increases in the synthesis of nucleic acids, cell membrane phospholipids, protein glycosylation, one-carbon metabolism as well as lipid metabolism.

CHAPTER 4

Relation between Energy Balance and Metabolic Profiles in Plasma and Milk of Dairy Cows in Early Lactation

Wei Xu[1,2], Jacques Vervoort[2], Edoardo Saccenti[3], Bas Kemp[1], Renny van Hoeij[1] & Ariette van Knegsel[1]

1. Adaptation Physiology Group, Department of Animal Sciences, Wageningen University & Research, the Netherlands

2. Laboratory of Biochemistry, Wageningen University & Research, the Netherlands

3. Laboratory of Systems and Synthetic Biology, Wageningen University & Research, the Netherlands

Submitted

4.1. Abstract

Alteration of metabolic status of dairy cows in early lactation is related to negative energy balance (NEB). In our study we show that metabolomics data can provide detailed information about energy balance and metabolic status. Little is known, however, about the relation between the metabolic profile of blood plasma and milk samples of dairy cows in NEB in early lactation. In this study, our aims were, *i*) to reveal metabolic profiles of plasma and milk samples by integrating results of nuclear magnetic resonance (NMR) and liquid chromatography triple quadrupole mass spectrometry (LC-QQQ-MS); *ii*) to investigate the relationship between energy balance and metabolic profiles of plasma and milk samples. For this study 24 individual dairy cows were studied in lactation week 2. Body weight, feed intake, and milk yield were monitored daily. Milk composition (fat, protein, and lactose) and net energy balance were calculated on a weekly basis. Plasma and milk samples were collected weekly and analysed through NMR and LC-QQQ-MS. From all plasma metabolites measured, 27 metabolites in plasma were correlated ($P < 0.05$) with energy balance. These plasma metabolites are related to body reserve mobilization from body fat, muscle, and bone, increased blood flow, and gluconeogenesis. From all milk metabolites measured, 30 metabolites in milk were correlated ($P < 0.05$) with energy balance. These milk metabolites are related to cell apoptosis and cell proliferation. The metabolites detected in plasma and milk are to a limited extent interrelated, 9 metabolites in milk and plasma were correlated to each other and were correlated ($P < 0.05$) with energy balance. These metabolites are mainly related to hyperketonemia, β-oxidation of fatty acids, and one carbon metabolism either in the body, or in the mammary gland, or in both. In conclusion, metabolic profiles of plasma and milk clearly reflect the metabolism in the body or in the mammary gland of dairy cows in NEB in early lactation.

4.2. Introduction

In early lactation of dairy cows, the elevated energy requirements for milk production combined with a relatively low energy intake can result in a negative energy balance (NEB) [27,28]. A severe NEB is associated with a greater risk of metabolic disorders [115], and compromised health and fertility [8,116]. To compensate for the energy deficit in early lactation, dairy cows mobilize body reserves [8], such as, body fat and muscle protein [194]. Traditionally, the concentration of metabolites in plasma and in milk have been used to diagnose a NEB status and associated metabolic disorders. For example, changed profiles of free fatty acids in plasma are related to a NEB status [42], and BHB in plasma or in milk is a biomarker for subclinical ketosis in dairy cows [99,195].

In the past decade, the combination of metabolomics techniques and advanced statistical methods makes it possible to identify and quantify low abundant molecules from bio-fluids. Using an integrated analysis complex correlation matrices can be obtained by combining results from NMR and LC-MS measurements with complementary data from plasma, milk, or urine [170,196]. In dairy cows, Klein *et al.* (2013) reported the detection of amino acids and carboxylic acids in plasma and milk using gas chromatography mass spectrometry and NMR and discussed only the metabolites related to ketosis [143]. Maher *et al.* (2013) reported that the majority of milk metabolites were not correlated to their concentration in plasma [170]. The biomarkers for heat stress of cows were identified through the spectra of LC-MS and NMR in plasma [57]. To our knowledge, the correlation between metabolic profiles of plasma and milk samples has only been studied to a limited extend [143,170]. In addition, the differences or similarities in metabolic profiles of cows based on plasma and milk samples that are associated with energy balance have not been studied systematically.

In the current study, we hypothesize that metabolic profiles of cows based on plasma and milk might differ concerning their relation with energy balance. Aims of this study were to reveal metabolic profiles of cows based on plasma and milk samples by combining results of NMR and LC-QQQ-MS and to investigate the relation between energy balance and metabolic profiles of cows based on plasma and milk samples.

4.3. Materials and Methods

4.3.1. Animals and sample collection

The experimental protocol of the study was approved by the Institutional Animal Care and Use Committee of Wageningen University and was conducted at Dairy Campus research farm (WUR Livestock Research, Lelystad, the Netherlands). The experimental design was described previously [126]. In the current study, we focus on cows with 100 percent energy level, and parity 2 or 3, as well as cows come from same batch. Earlier we reported the relation between milk metabolites identified through LC-QQQ-MS and energy balance of 31 dairy cows in lactation week 2 and 7 (Chapter 2), the LC-QQQ-MS data obtained from milk samples of 24 cows in lactation week 2 were used again in the current study. These 24 dairy cows were randomly assigned to one of two DPL (0 or 30 days) before calving. Prepartum, cows with a 0 day DPL received a lactation ration based on grass silage and corn silage (6.4 MJ net energy for lactation (NE)/kg dry matter). Cows with a 30 days DPL received a dry cow ration based on grass silage, corn silage and wheat straw (5.4 MJ NE/kg DM). Postpartum, all cows received the same basal lactation ration as provided to lactating cows prepartum plus additional concentrates. Postpartum, concentrate supply increased stepwise with 0.3 kg/d till 8.5 kg/d on 28 DIM. Body weight, milk yield, and feed intake were recorded daily. During lactation, cows were milked twice daily at ~0600 hours and ~1800 hours.

Milk samples for analysis of fat, protein and lactose percentage (ISO 9622, Qlip, Zutphen, the Netherlands) were collected four times per week (Tuesday afternoon, Wednesday morning, Wednesday afternoon, and Thursday morning). Milk samples were analysed as a pooled sample per cow per week and used to calculate average fat, protein and lactose yield per week. Plasma samples for metabolomics studies were collected on Thursday after the morning milking, between 3 and 1 hours before the morning feeding. Milk samples for metabolomics studies were collected at Friday morning each week. All samples were collected and stored at -20 °C until analysis. Milk production traits were averaged per week.

4.3.2. Energy intake and energy balance

Roughage and concentrate were supplied separately and daily intakes were recorded per individual cow using roughage intake control troughs (Insentec, Marknesse, the Netherlands). Energy balance was calculated per week according to the Dutch net energy evaluation (VEM) system, as the difference between energy intake and the estimated net energy requirements for maintenance and milk yield [33,168]. According to the VEM system, the daily requirement for maintenance is 42.4 VEM/kg$^{0.75}$ of BW, the requirement for milk yield is 442 VEM/kg of fat- and protein-corrected milk. Energy intake and energy balance are expressed in kJ/kg$^{0.75}$·d, where kg$^{0.75}$ indicates metabolic body size [33].

4.3.3. Sample preparation

Sample preparation using in LC-MS and NMR was performed as described earlier by [66]. Briefly, the fat layer of milk was removed by addition of deuterated chloroform and subsequent centrifugation (12,000 rpm, 15 min, Centrifuge 5424, Eppendorf). Blood plasma and milk samples were first thawed to room temperature. Subsequently 175 µL of milk serum or blood plasma was mixed with 175 µL of phosphate buffer (pH = 7.0) and these samples were subsequently filtered to remove protein using a Pall 0.5 mL 10-kDa cut-off spin filter (Millipore Corp., Billerica, MA) with centrifugation at 12,000 rpm for 15 min.

4.3.4. Measurement in LC-QQQ-MS and data pre-processing

Quantification of metabolites was described previously (Chapter 2). The fat layer of milk was removed by addition of deuterated chloroform and subsequent centrifugation (12,000 rpm, 15 min, Centrifuge 5424, Eppendorf). Measurements were performed with a triple quadrupole mass spectrometer (Shimadzu LC-QQQ-MS; LCMS-8040) using the PFPP method as described earlier [127,128]. The sample injection volume used was 1 µL, and a single analysis took 25 minutes. From LC-QQQ-MS spectra, metabolites were regarded as reliably identified, when more than 60% observations in all samples showed a reliable value for a certain metabolite.

4.3.5. Measurement in NMR and data pre-processing

Prepared samples were measured using a Samples were measured using a Bruker NMR spectrometer Avance III with a 600 MHz/54 mm UltraShielded Plus magnet equipped with a CryoPlatform cryogenic cooling system, a BCU-05 cooling unit and an ATM automatic tuning and matching unit. Measurements were done at 300 K. 1D nuclear Overhauser enhancement spectroscopy (NOESY) spectra were obtained. Baseline corrections, alignment, and calibration to internal maleic acid was done for all spectra.

Assignment of metabolites resonances was performed using published literature, the Human Metabolome Database version 2.0 online library (http://hmdb.ca/), as well as internal standards.

4.3.6. Integrated analyses and software

If a metabolite could be identified by LC-MS as well as NMR, the intensity of this metabolite was quantified based on the NMR spectra. NMR is a very reliable technique with high reproducibility. Integrated analysis was applied as described earlier [166,170]. Briefly, Pearson correlation was applied to any two columns in a matrix. Pearson correlation coefficient (*r*) and corresponding *P* value were obtained by function "cor.test ()" in R environment (version 3.3.3).

4.4. Results

4.4.1. Milk production and energy balance

In lactation week 2, average (SD in parentheses) BW of 24 dairy cows was 627.4 (56.4) kg, milk yield was 28.1 (6.7) kg/d, fat- and protein-corrected milk production (FPCM) was 32.3 (7.3) kg/d, milk fat yield was 1.4 (0.4) kg/d, milk protein yield was 1.1 (0.2) kg/d, and milk lactose yield was 1.3 (0.3) kg/d. Energy balance of dairy cows was -180.0 (219.1) kJ/kg$^{0.75}$·d, which was negatively correlated with milk yield and milk production traits, *r* ranged from -0.78 to -0.91 (Fig. 4.1).

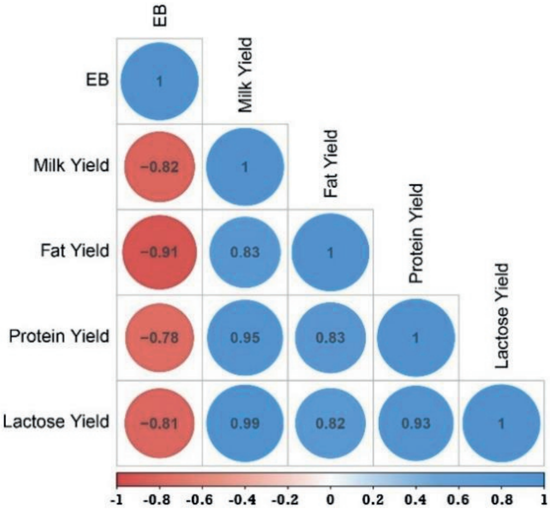


Fig. 4.1. The correlation of energy balance (EB) with milk yield, fat- and protein-corrected milk production (FPCM), milk fat yield, milk protein yield, and milk lactose yield.

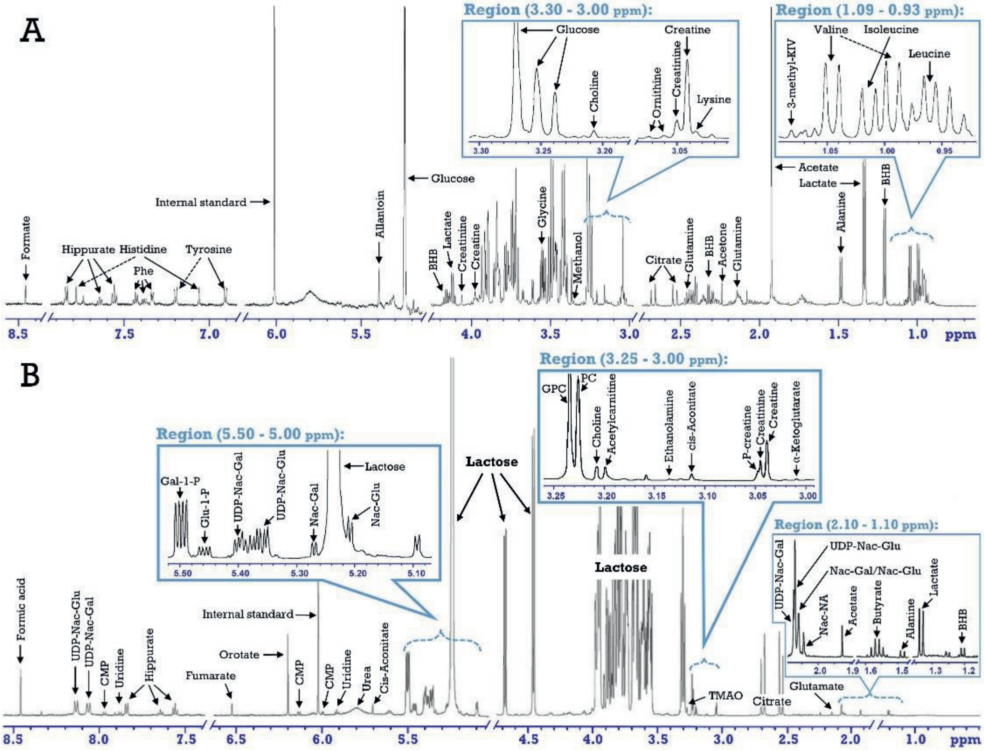


Fig. 4.2. Metabolites identified from nuclear magnetic resonance spectra in plasma (A) and milk (B) of dairy cows in early lactation. Metabolites correlated with energy balance are labelled with red.

Abbreviations: BHB, β-hydroxybutyrate; CMP, cytidine monophosphate; Gal-1-P, galactose-1-phosphate; Glu-1-P, glucose-1-phosphate; GPC, glycerophosphocholine; Nac-Gal, N-acetyl-glucosamine; Nac-Glu, N-acetyl-glucosamine; Nac-Na, N-acetyl-neuraminic acid; PC, phosphocholine; P-creatine, phosphocreatine; phosphocholine; TMAO, trimethylamine N-oxide; UDP-Nac-Gal, uridine diphosphate-N-acetyl-galactosamine; UDP-Nac-Glu, uridine diphosphate-N-acetyl-glucosamine.

4.4.2. Measurement by LC-MS and NMR and integrated analysis

Plasma and milk samples of 24 dairy cows in lactation week 2 were measured using both LC-QQQ-MS and NMR. In LC-QQQ-MS spectra, 97 metabolites were initially targeted for both plasma and milk samples, and in LC-MS spectra of plasma 43 metabolites and of milk 41 metabolites could be reliably detected (*Appendix* Table 3.1 and 4.1). In the NMR spectra (0.01 ppm), bins which were correlated to energy balance were selected and the corresponding NMR resonances (peaks) were specifically integrated by carefully selecting peaks that did not show overlap in the NMR spectra (*Appendix* Table 3.1 and 4.1). Through the combination of LC-MS and NMR (Fig. 4.2), a total of 53 and 65 metabolites were identified from plasma and milk samples, respectively. Of these metabolites, 38 metabolites were detected in both blood plasma and milk. In NMR spectra of blood, isobutyrate 1,2-propanediol, α-ketoisovalerate were tentatively labelled in NMR spectra of blood. In milk β-alanine was tentatively assigned. Detailed information of these metabolites in blood plasma and milk is presented in *Appendix* Table 3.1 and 4.1, respectively.

4.4.3. *Relation of energy balance with metabolic profiles in plasma and in milk*

Energy balance was correlated with several metabolites either in milk, or in plasma, or in both ($P < 0.05$). In total 9 detected metabolites were correlated with energy balance both in plasma and in milk, r ranged from -0.42 to -0.80 and from 0.43 to 0.59 (Fig. 4.3-A). Many of these 9 metabolites showed a correlation between their concentrations in plasma and in milk, with r ranging from 0.49 to 0.92. Of the 53 detected and identified metabolites in plasma, 24 were correlated with energy balance of dairy cows (Fig. 4.3-A, 4.3-B and 4.3-D), r ranging from -0.80 to -0.43, and from 0.43 to 0.84. Of the 65 detected and identified metabolites in milk, 30 were correlated with energy balance of dairy cows (Fig. 4.3-A, 4.3-C and 4.3-D), r ranging from -0.79 to -0.42, and from 0.43 to 0.71. In total 10 and 12 metabolites correlated with energy balance that were uniquely detected from plasma and milk samples, respectively (Fig. 4.3-D).

Based on their correlation with energy balance, metabolites were grouped as one of three groups, *i*), energy balance correlated with metabolites both in plasma and in milk (Fig. 4.3-A); *ii*), energy balance correlated with metabolites only in plasma (Fig. 4.3-B and 4.3-D); *iii*), energy balance correlated with metabolites only in milk (Fig. 4.3-C and 4.3-D).

4.5. Discussion

4.5.1. *Metabolites in plasma and in milk related to energy balance*

In the current study, 9 metabolites detected in both plasma and milk were related to energy balance of dairy cows in early lactation and had a correlation between their concentration in plasma and milk. These metabolites were acetone, acetyl-carnitine, aspartate, BHB, carnitine, creatinine, glycine, hydroxyproline, and thymidine.

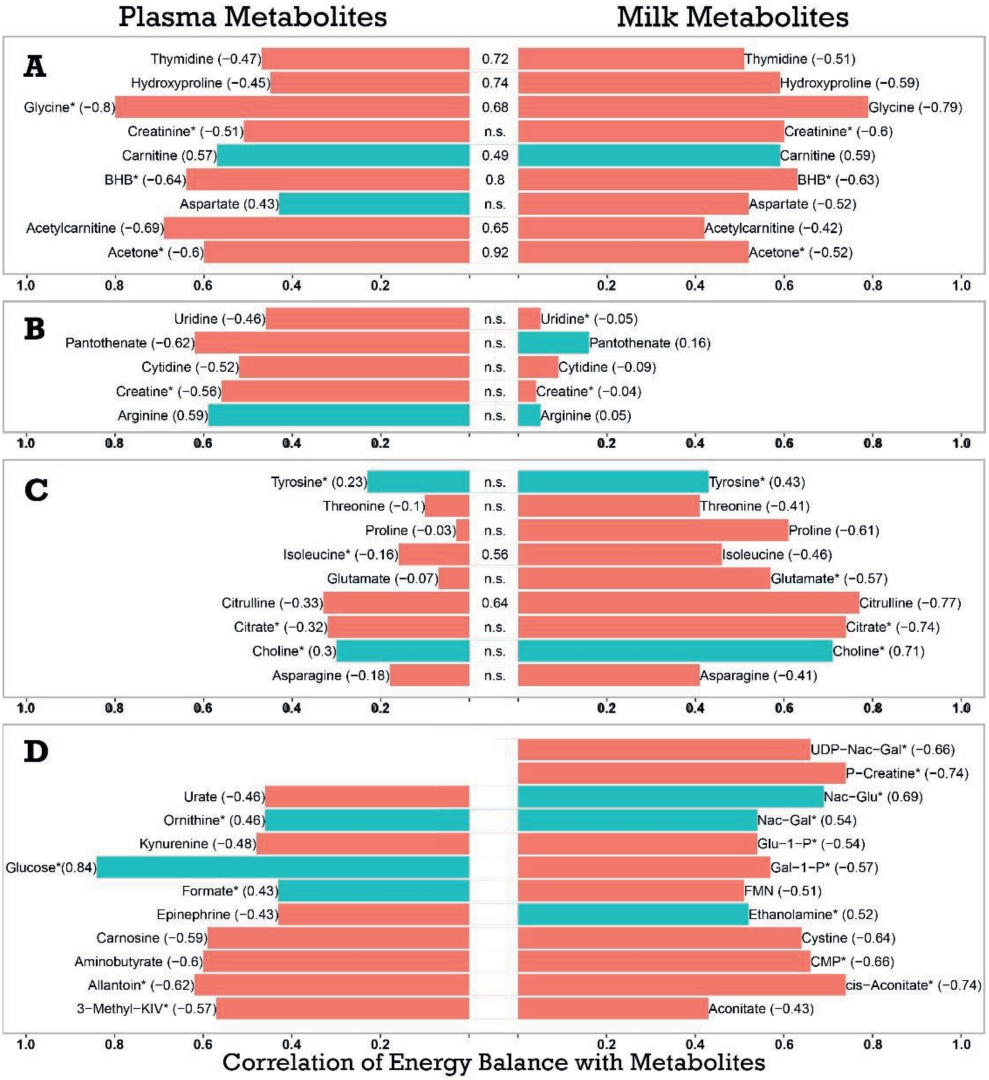


Fig. 4.3. Pearson correlations of energy balance with metabolites in lactation week 2. A, metabolites in both plasma and milk present a correlation with energy balance; B, metabolites in plasma, but not in milk present a correlation with energy balance; C, metabolites in milk, but not in plasma with a correlation with energy balance; D, unique metabolites in plasma or in milk present a correlation with energy balance. Cyan bar presents a positive correlation; red bar presents a negative correlation. Metabolites detected by NMR are marked by an asterisk. Value within brackets is coefficient of Pearson correlations. Value in middle of milk metabolites and plasma metabolites is r of metabolites between its concentration in plasma and in milk.

Abbreviations: BHB, β-hydroxybutyrate; cAMP, 3',5'-cyclic adenosine monophosphate; CMP, cytidine monophosphate; FMN, Flavin mononucleotide; Glu-1-P, glucose-1-phosphate; Gal-1-P, galactose-1-phosphate; GPC, glycerophosphocholine; Nac-Gal, N-acetyl-galactose; Nac-Glu, N-acetyl-glucosamine; 3-methyl-KIV, 3-methyl-2-oxovaleric acid; UDP-Nac-Gal, uridine diphosphate-N-acetyl-galactosamine; UDP-Nac-Glu, uridine diphosphate-N-acetyl-glucosamine.

Of all detected amino acids and their derivatives, glycine presents the most striking correlation with energy balance in both plasma (r = -0.80) and milk (r = -0.79). This was consistent with previous studies where dairy cows had an increased glycine concentration in plasma [197] and in milk [54] around calving. In dairy cows in NEB, an increased glycine concentration in plasma could be explained by the breakdown of muscle protein [197], or alternatively the *de novo* synthesis of glycine from threonine and serine was also suggested [155]. In our study, glycine in plasma had a positive correlation with its concentration in milk (r = 0.68). Plasma glycine, can be absorbed by the mammary gland of dairy cows [198]. Based on the results of our study we propose that the increased concentration of glycine in milk is due to one carbon processes in which choline is converted into glycine. Choline can provide methyl groups for the process of cell proliferation in the mammary gland [199]. Glycine in both plasma and milk (Chapter 2) has the potential to be used as an indicator for the energy balance and metabolic status of dairy cows [200]. As will be discussed in more detail below, we propose that the increased concentration of glycine in both plasma and milk is due an increase in one carbon metabolic processes in which choline is converted into glycine. Choline can provide methyl groups for the process of cell proliferation in the mammary gland [199]. In our study, energy balance was negatively correlated with Gly/Ala in plasma (r = -0.72) and in milk (r = -0.82), the correlation between Gly/Ala in plasma and Gly/Ala in milk was 0.85. The ratio of glycine to alanine (Gly/Ala) in plasma was used as a biomarker for mal-nutritional status of dairy cows in early lactation [200]. However, alanine itself was not correlated to energy balance ($P < 0.05$), probably because of the dual nature of alanine in the glucose-alanine cycle in muscle and hepatocytes. Based on our observations, we conclude that the previously proposed Gly/Ala ratio is more dependent on the fluctuation of the glycine levels than of the changes in the concentration of alanine and that monitoring glycine levels is more indicative of changes in energy balance than monitoring the Gly/Ala ratio.

Energy balance was negatively correlated with thymidine in both plasma (r = -0.47) and milk (r = -0.51). Thymidine plays an important role in DNA synthesis [201]. The thymidine in the mammary gland can be speculated to originate from plasma, which is supported by the positive correlation (r = 0.72) between thymidine in plasma and in milk.

Energy balance was negatively correlated with hydroxyproline (HYP) in both plasma (r = -0.45) and milk (r = -0.59). In early lactation, dairy cows are suffering from low calcium and magnesium levels [202]. Therefore, dairy cows mobilize minerals from bone to maintain the mineral balance in plasma [202]. Released from bone degradation, HYP in plasma could indicate the balance between bone formation and degradation [203]. Moreover, the high concentration of HYP in milk of dairy cows in NEB indicates that HYP could be used for cell renewal in the mammary gland. During cell proliferation, HYP and proline are the major amino acids in collagen that is the main structural protein in connective tissues [204]. Both HYP and proline detected were present in high amounts in milk of cows in NEB ($r = -0.61$). The high concentration of hydroxyproline and proline in milk of cows in NEB indicates that the compounds could be used for forming connective tissues for cell renewal, possibly also in the uterus, rumen, and muscle. Epithelial cells of mammary gland are permeable to plasma HYP, which is supported by the strong positive correlation ($r = 0.74$) between HYP in plasma and in milk.

Energy balance was negatively correlated with acetone and BHB in both plasma and milk. In addition, the correlation between acetone in plasma and in milk was 0.92, and between BHB in plasma and in milk was 0.80. In early lactation, the high requirement for glucose coincides with high plasma concentrations of ketone bodies [205]. Plasma ketone bodies (acetone and BHB) are well known to be related to the incomplete β-oxidation of mobilized body fat [150,206], resulting in ketosis. Acetone and BHB, either in plasma, or in milk, were used to diagnose ketosis or subclinical ketosis of cows in early lactation [207,208], which is line with our result that energy balance is correlated with ketone bodies in both plasma and milk. In plasma, acetone and BHB could be used not only for energy supply [209], but also for the synthesis of short- and milk medium-chain fatty acids in the mammary gland [77]. Results of current study could indicate a direct relation between ketone bodies in plasma and in milk, which is supported by the arteriovenous difference of BHB and acetoacetate plus acetate in the mammary gland [210]. This confirms the value of milk ketone body levels as an indicator for incomplete oxidation of fatty acids and ketosis at systemic level.

The correlation of carnitine in plasma with its concentration in milk was 0.49, and the correlation of acetyl-carnitine in plasma with its concentration in milk was 0.65, which is suggested to be related to the role of carnitine and acetyl-carnitine in fat metabolism [157]. Carnitine and acetyl-carnitine play important roles to transport activated long-chain fatty acids from cytosol into mitochondria, which happens both in liver and in the mammary gland [211]. The correlation of carnitine, acetyl-carnitine, BHB, and acetone in plasma with their concentration in milk, and their positive correlation with energy balance indicates the oxidation of fatty acids in the liver, and the use of fatty acids as energy source and for milk fat synthesis in the mammary gland of dairy cows in NEB status.

Energy balance was negatively correlated with citrate in milk ($r = -0.74$). Citrate not only plays a role in the TCA cycle, but also forms one of the main products in buffer systems in milk [191]. In our previous study, milk citrate was an indicator of energy balance of dairy cows (Chapter 2). In the mammary gland, epithelial cells are impermeable to citrate [191], we thus speculate that the concentration of citrate in milk can reflect the fat metabolism in the mammary gland itself.

4.5.2. Metabolites in plasma related to energy balance

In the current study, a number of plasma metabolites were related to energy balance, but no such relationship was observed in milk samples. These plasma metabolites were arginine, pantothenate, allantoin, glucose, and epinephrine.

Energy balance was positively correlated with arginine in plasma ($r = 0.59$), but not in milk ($P > 0.05$). In the urea cycle (Fig. 4.4), arginine metabolism is not only related to nitrogen metabolism but also related to nitric oxide (NO) production. Nitric oxide is a major vasodilator factor [212]. The low level of arginine in dairy cows in NEB indicates that NO in plasma increases with a concomitant increase in blood flow for an increased supply of nutrients to support milk production in the mammary gland. The increased blood flow also facilitates the absorption of nutrients from the small intestine [213]. In the current study, energy balance was negatively correlated with plasma pantothenate ($r = -0.62$) and plasma allantoin ($r = -0.70$). Both pantothenate and allantoin are nutrients for dairy cows, which are derived from feed sources or produced by ruminal microbes [117,214]. Further identification of the plasma metabolites from rumen microbes may give insights into rumen function and its effect on milk composition of dairy cows.

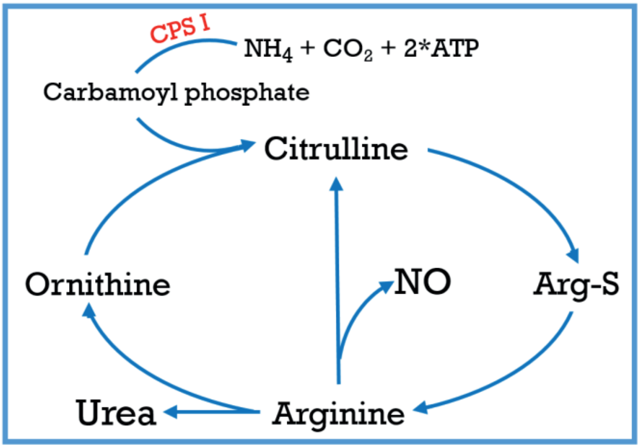


Fig. 4.4. The nitric oxide (NO) production from the urea cycle, and the association between urea cycle and tricarboxylic acid (TCA) cycle.

Abbreviations: Arg-S, argininosuccinate; ATP, adenosine triphosphate; CPS I, carbamoyl phosphate synthetase I.

Energy balance was positively correlated with plasma glucose ($r = 0.80$), which is in line with earlier studies that plasma glucose decreases after parturition [215,216]. Dairy cows use glucose not only as an energy source [217], but also as the precursor to synthesize lactose in milk [218] and as methyl donor via one-carbon metabolism in DNA and histone methylation during cell renewal in the mammary gland [155,219]. Plasma glucose concentration is regulated by insulin that is low for dairy cows in NEB [126]. A low insulin concentration decreases the uptake of glucose by peripheral tissue and thus facilitates uptake of glucose by the mammary gland which has insulin-independent glucose transporters [220]. In the mammary gland, one molecule glucose is converted to galactose, which is combined with another molecule of glucose to synthesize lactose [172]. In our study, low glucose concentration in plasma during NEB could mirror the high priority of mammary gland for glucose [160,221].

The shortage of glucose could be compensated by gluconeogenesis. For example, branched-chain amino acids (BCAAs, isoleucine, leucine, and valine) can be used for gluconeogenesis in liver [222]. In the current study, however, none of BCAAs in plasma was correlated with energy balance in the current study ($P < 0.05$).

Body fat metabolism accounts for the most of energy lost or gained of dairy cows [223], however, muscle protein is also mobilized [223,224]. In plasma, energy balance was negatively correlated with plasma creatine ($r = -0.57$) and plasma creatinine ($r = -0.76$), which indicates that energy balance is associated with the mobilization of muscle protein in dairy cows in NEB. Our results are in line with a previous study that the mobilized muscle protein is related to the greater plasma creatine and creatinine in cows with low energy balance [225]. Energy balance was correlated with creatinine both in plasma (r = -0.51) and in milk (r = -0.60). Energy balance was correlated with creatine only in plasma (r = -0.56), and the permeability of mammary epithelium cells to creatine is still unknown. In the process of protein mobilization, the released amino acids in plasma are intensively used for milk protein synthesis or gluconeogenesis where not all amino acids are used to the same extent [144].

Epinephrine in plasma is one of hormone presenting a correlation with energy balance (r = -0.43), which is consistent with that infusion of epinephrine could cause increased free fatty acids (FFA) and glucose in plasma [226,227]. The plasma glucose is low in dairy cows in NEB, which could be regarded as the hormonal regulation for ruminants during the adaptation period [226]. Besides its role to stimulate body fat mobilization in early lactation [11], epinephrine is known to stimulate gluconeogenesis in the liver thereby compensating the low levels of glucose to some extent [228]. In mammals, epinephrine could be synthesized via tyrosine in adrenal gland [229]. Tyrosine in plasma, however, was not correlated with energy balance ($P > 0.05$).

4.5.3. Metabolites in milk related to energy balance

Milk metabolites can originate from several sources, including being transferred from blood, being leaked from damaged somatic cells and bacteria present in milk, or being secreted from mammary epithelial cells [125,230]. The correlation between energy balance and metabolites in milk, rather than in plasma could indicate the metabolism in the mammary gland of dairy cows in lactation week 2.

Energy balance was negatively correlated with galactose-1-phosphate (Gal-1-P, r = -0.57) and glucose-1-phosphate (Glu-1-P, r = -0.54) in milk. Both Gal-1-P and Glu-1-P are two intracellular intermediates in lactose synthesis [134]. Great levels of Gal-1-P and Glu-1-P of dairy cows in NEB were suggested as the leakage of cellular content due to cell apoptosis in the mammary gland in early lactation [66]. Cell apoptosis happens in coordination with cell proliferation, which is indicated by the 65% increased total DNA used for meiosis around 10 days pre- and post-partum [180]. The increased nucleotides metabolism could also explain the positive correlation of energy balance with milk cytidine monophosphate (CMP, r = -0.66) and thymidine (r = -0.51). The positive correlation between thymidine in plasma and its concentration in milk (*r* = 0.72) indicates that the source of thymidine in the mammary gland is plasma. Besides the synthesis of nucleotides, the correlation of energy balance with milk ethanolamine (*r* = 0.52) and milk GPC (*r* = 0.40, *P* = 0.05), could be related to the process of cell membrane synthesis [182]. Further studies are expected to detect more correlation of energy balance with metabolites related to cell membrane synthesis, such as, phosphocholine (PC, *r* = -0.39, *P* = 0.06). In the context of cell membrane synthesis, the correlation of energy balance with detected acetyl-derivatives including N-acetyl-galactosamine (*r* = 0.54), N-acetyl-glucosamine (*r* = 0.69), and UDP-N-acetyl-galactosamine (*r* = -0.66), could be explained by their roles in the glycosylation of proteins.

If the ratio of GPC-to-PC in milk is less than 2.5, dairy cows have a great risk for developing ketosis [64]. In our study, energy balance was positively correlated with the ratio of GPC-to-PC (*r* = 0.49, *P* = 0.02), which indicates that dairy cows in NEB have an increased risk for ketosis. Therefore, the correlation between energy balance and metabolites in milk could indicate the apoptosis and proliferation of cell in the mammary gland of dairy cows in NEB.

Energy balance was correlated with choline in milk (*r* = 0.71), but not with choline in plasma. Choline acts as a methyl donor for biological processes involving the folate cycle, redox balance status and cell renewal [145], and choline is regarded as a limiting nutrient for transition dairy cows [147]. The supplementation of rumen-protected choline (RPC) can increase milk production [231,232], which is possible due to the facilitating effect of choline to export fat into plasma from the liver in cows [150,233]. Based on this hypothesis, dairy cows supplemented with RPC should have low plasma BHB and FFA. In our study, choline in plasma is not correlated with plasma BHB or acetone (*P* > 0.05), which is supported by studies that supplemented PRC is not correlated with a decreased plasma BHB and FFA of dairy cows [231,234]. We speculate that the majority of choline is consumed in the mammary gland to facilitate the cell proliferation in early lactation, which might be related to the start of a new lactation and increase in milk production.

4.5.4. Integrated analysis by combination of NMR and LC-MS

A series of metabolites in plasma and in milk, for example, plasma BHB and milk citrate, have been reported in previous studies using either mass spectrometry, or NMR, or both [54,125,235]. A total of 12 metabolites (either in plasma, or in milk) were detected both in NMR and in LC-MS. These metabolites showed a correlation *r* more than 0.85 between NMR and LC-QQQ-MS measurements, in plasma, alanine (*r* = 0.95), choline (*r* = 0.87), creatine (*r* = 0.87), creatinine (*r* = 0.85), glutamine (*r* = 0.85), glycine (*r* = 0.96), lactate (*r* = 0.95), and valine (*r* = 0.95); in milk, acetyl-carnitine (*r* = 0.97), choline (*r* = 0.98), CMP (*r* = 0.99), and uridine (*r* = 0.97). The high correlations of intensity of these metabolites between NMR and LC-MS indicated that both NMR and LC-QQQ-MS measurements were performed with the desired quality and that the datasets were reliably interpreted.

In general, to our knowledge, our study is the first that reports the metabolic profiles in plasma and in milk, and their correlation with energy balance of dairy cows in early lactation. Of all metabolites related to energy balance, the majority of metabolites either in plasma, or in milk, could be related to the metabolic pathways either in body, or in the mammary gland, or in both.

4.6. Conclusion

In this study, we revealed metabolic profiles in plasma and milk of 24 dairy cows in lactation week 2 through the combined measurement of NMR and LC-MS. A total 53 and 65 metabolites were identified in plasma and in milk, respectively. In total, 27 metabolites in plasma were correlated with energy balance. These plasma metabolites are related to body reserve mobilization from body fat, skeleton muscle, and bone, increased blood flow, and gluconeogenesis. In total, 30 metabolites in milk were correlated with energy balance. These milk metabolites are related to the apoptosis and proliferation of cells in the mammary gland. In total, 9 metabolites were related to energy balance both in plasma and in milk. These metabolites can reflect the metabolic pathways of dairy cows either in body, or in the mammary gland, or in both, which are mainly related to the hyperketonemia, β-oxidation of fatty acids, and one carbon metabolism during cell proliferation in early lactation.

CHAPTER 5

A Preliminary Study on Milk Proteomics of Dairy Cows in Early Lactation

Wei Xu[1,2], Sjef Boeren[2], Bas Kemp[1], Ariette van Knegsel[1] & Jacques Vervoort[2]

1. Adaptation Physiology Group, Department of Animal Sciences, Wageningen University & Research, the Netherlands

2. Laboratory of Biochemistry, Wageningen University & Research, the Netherlands

5.1. Abstract

Due to increased milk production in early lactation dairy cows normally have a negative energy balance (NEB). In our metabolomics studies as discussed in this thesis, we observed that in the mammary gland of dairy cows in NEB both cell apoptosis and cell renewal are paramount. As a consequence of these cellular processes differences in milk protein composition as well as differences in post translational modifications of milk proteins can be expected to occur. In a bottom-up proteomics study using a nano-LC-Orbitrap FT-MS system, we focused on differences in post translational modifications with special emphasis on protein methylation as well as differences in protein concentration. Immunoglobulin and β-lactoglobulin had a decreased methylation level for cows in NEB, supporting the results obtained from our metabolomics studies that mammary gland cells of cows in NEB have a lack of methyl-donors. Legumain, α-2-HS-glycoprotein, vitamin D binding protein, glycoprotein 2, folate receptor alpha, peptidyl-prolyl cis-trans isomerase B and apolipoprotein A-IV were upregulated in milk of cows in NEB. Two proteins, IGK and an immunoglobulin-like protein, were downregulated in NEB.

5.2. Introduction

In early lactation of dairy cows, elevated energy requirements for milk production result in an energy deficit or negative energy balance (NEB) [27,28]. A severe NEB is related to a great risk of metabolic disorders and related diseases, such as fatty liver and ketosis [158,159]. Ketosis in dairy cows was observed to be related to the concentration of metabolites in plasma or in milk [64]. In our studies we could relate energy balance to the alteration of metabolic profiles both in plasma and in milk (Chapter 2, 3, and 4). There are very few milk proteomics studies related to cows in NEB [66]. An improved proteomic profile of milk could result in a better understanding of biological processes involved in negative energy balance.

For the past decades, proteomics techniques have identified and characterized a numerous amount of proteins in milk [69,236,237]. Proteins present in milk have been related to functions important for the immune and digestive systems of calves [238] as well as for differences between Holstein and Jersey breeds of dairy cows [236]. There have been limited studies on studies on post-translational modifications of proteins in milk and most of these studies are focusing on glycosylation patterns. Glycosylation patterns of milk proteins are considered important for the stability and the biological function of the proteins involved. Based on our metabolomics studies as discussed chapter 2, 3 and 4 we expect changes to occur in important metabolic pathways in cells in the mammary gland in cows in early lactation. One of the pathways directly related to NEB was the folate cycle, a metabolic process related to DNA/RNA synthesis and redox control. The amount of choline in milk was decreased in cows in NEB. Choline is the major methyl donor for the one carbon cycle and we expect that differences in milk protein methylation might occur in relation to energy balance of the dairy cows. Protein methylation is a posttranslational modification that occurs predominantly on arginine and lysine residues [239,240].

The objective of this (preliminary) study is to apply advanced proteomic techniques to reveal differences in methylation of milk proteins and to reveal differences in protein composition. Based on this study we could establish a relationship between energy balance and the methylation status of some important proteins in milk.

5.3. Materials and Methods

5.3.1. Animals and energy balance

The experimental protocol for the study was approved by the Institutional Animal Care and Use Committee of Wageningen University and was conducted at Dairy Campus research farm (WUR Livestock Research, Lelystad, the Netherlands). The experimental design was described previously [87]. In Chapter 2, we reported the correlation between milk metabolites identified through LC-MS and energy balance of 31 dairy cows in lactation week 2, of these 31 cows milk samples of 5 cows were used in the current study. Energy balance of individual dairy cows was calculated as discussed before in Chapter 2.

5.3.2. Milk serum separation

Milk samples were centrifuged at 1,500 g for 10 min (Eppendorf centrifuge). The pellet was removed and the obtained supernatant was ultra-centrifuged at 100,000 g for 90 min at 30 °C (Beckman L-60, rotor 70Ti). After ultracentrifugation, samples were separated into three phases. The top layer was milk fat, the middle layer was milk serum, and the bottom layer was casein. Milk serum was used for the protein concentration determination (BCA assay) and filter aided sample preparation (FASP) as described below.

5.3.3. BCA assay

Milk protein concentration was determined by BCA Protein Assay Kit 23225 (Thermo Scientific Pierce), according to the manufacturer's instructions. Bovine serum albumin (BSA) was used as standard for making a calibration curve. Subsequently, the milk serum protein concentration was determined, based on the standard curve, which covers the protein concentration from 0.02–2μg/μL.

5.3.4. Filter aided sample preparation

The methods of proteomics analysis are based on two previous articles [238,241]. Milk serum samples (20 μL) were diluted in SDT-lysis buffer (100 mM Tris/HCl pH 8.0+0.1 M Dithiotreitol) to get a 1 μg/μL protein solution. Then, 20 μL of sample was transferred to a low binding Eppendorf tube with 180 μL 100mM Tris/HCl pH 8.0+8 M urea containing acrylamide. The low binding Eppendorf tube was then incubated for 10 min while mildly shaking at room temperature. All of the sample was transferred to a Pall 3K Omega filter (10-20 kDa cutoff, OD003C34; Pall, Washington, NY, USA) without touching the filters poly-propylene side, and centrifuged at 12,000 rpm for 30 min. To the sample 100 µL 50 mM NH4HCO3 was added. Digestion was done overnight by adding trypsin while mildly shaking at room temperature. Finally, the sample was centrifuged at 12,000 rpm for 30 min, and 3.5 µL 10% trifluoroacetic acid (TFA) was added to the filtrate to adjust the pH value of the sample to around 2.

5.3.5. Measurement in LC-MS/MS and data analyses

Trypsin digested milk serum samples (18 µL) was injected on a 0.10*30 mm Magic C18AQ 200A 5 µm beads (Bruker Nederland B.V.) pre-concentration column (prepared in house) at a maximum pressure of 800 bar. Peptides were eluted from the pre-concentration column onto a 0.10*200 mm ReproSil-Pur 120 C18-AQ 1.9 µm beads analytical column with an acetonitrile gradient at a flow of 0.5 µL/min, using a gradient elution from 9% to 34% acetonitrile in water with 0.1 v/v% formic acid in 50 min. The column was washed using an increase in the percentage acetonitrile to 80% (with 20% water and 0.1 v/v% formic acid in the acetonitrile and the water) in 3 min. An Upchurch micro-cross was positioned between the pre-concentration and analytical column. An electrospray potential of 3.5 kV was applied directly to the eluent via a stainless-steel needle fitted into the waste line of the micro-cross. Full scan positive mode FTMS spectra in LTQ-Orbitrap XL (Thermo electron, San Jose, CA, USA) were measured between m/z of 380 and 1400. CID fragmented MSMS scans of the four most abundant 2 and 3+ charged peaks in the FTMS scan were recorded in data-dependent mode in the linear trap (MSMS threshold = 5,000).

Each run with all MSMS spectra obtained was analysed with Maxquant with Andromeda search engine[242]. Acrylamide modification of cysteine was set as fixed modification. Oxidation of methionine, N-terminal acetylation, de-amidation of asparagine or glutamine and methylation of lysine or arginine were set as variable modification for both identification and quantification. The bovine reference database for peptides and protein searches was downloaded as FASTA files from Uniprot (https://www.uniprot.org/) with reverse sequences generated by Maxquant. A set of 31 protein sequences of common contaminants was added including Trypsin (P00760, bovine), Trypsin (P00761, porcine), Keratin K22E (P35908, human), Keratin K1C9

(P35527, human), Keratin K2C1 (P04264, human), and Keratin K1C1 (P35527, human). A maximum of two missed cleavages were allowed and mass deviation of 0.5 Da was set as limitation for MS/MS peaks and maximally 4.5 ppm deviation on the peptide m/z during the main search. The length of peptides was set to at least seven amino acids. Finally, proteins were displayed based on minimally 2 distinct peptides of which at least one unique.

The function of the identified proteins was checked in the UniprotKB database released (http://www.uniprot.org/). To select the milk proteins that significantly correlated with energy balance, a Pearson correlation was calculated using the Log LFQ intensities from the MaxQuant filtered protein groups output table.

5.4. Results and Discussion

In lactation week 2, the average body weight of the 5 dairy cows was 640.1 (56.4) kg, milk yield was 28.1 (6.7) kg/d, FPCM was 32.3 (7.3) kg/d, milk fat yield was 1.4 (0.4) kg/d, milk protein yield was 1.1 (0.2) kg/d, and milk lactose yield was 1.3 (0.3) kg/d. The energy balance of 5 dairy cows was -326.7, -49.6, 80.4, 102.8 and 177.2 kJ/kg$^{0.75}$·d, respectively.

The protein concentration in milk serum of dairy cow 1, 2, 3, 4, and 5 was 1.0, 1.3, 1.9, 1.4, and 1.3 μg/mL, respectively.

A total of 219 proteins were identified and quantified after strict filtering (see materials and methods). From these 219 proteins, the methylation sites of 32 proteins could be identified and quantified. Lysine or arginine methylation is a well-known post-translational modification found on both nuclear and cytoplasmic proteins [243,244]. The methylation of lysine and arginine residues is catalysed by methyltransferases. For protein methylation processes S-adenosylmethionine is required. Proteins that are methylated are involved in a number of different cellular processes, including transcriptional regulation, RNA metabolism and DNA damage repair [245]. In dairy cattle in NEB there is a heavy demand for S-adenosylmethionine for processes related to DNA/RNA synthesis and redox balance. In our metabolomics studies we reported a choline deficiency in cows in NEB. We expected that this choline deficiency could be reflected in post translational modifications of proteins. Of these 32 proteins, 2 milk proteins, immunoglobulin and β-lactoglobulin were observed to have post-translational methylation differences that were related to energy balance (Fig. 5.1). Immunoglobulin and β-lactoglobulin are two abundant proteins in milk. Differences in methylation status of lower abundant proteins could not be established with certainty and will require more extensive studies with either higher concentrations of proteins or more sensitive instrumentation.

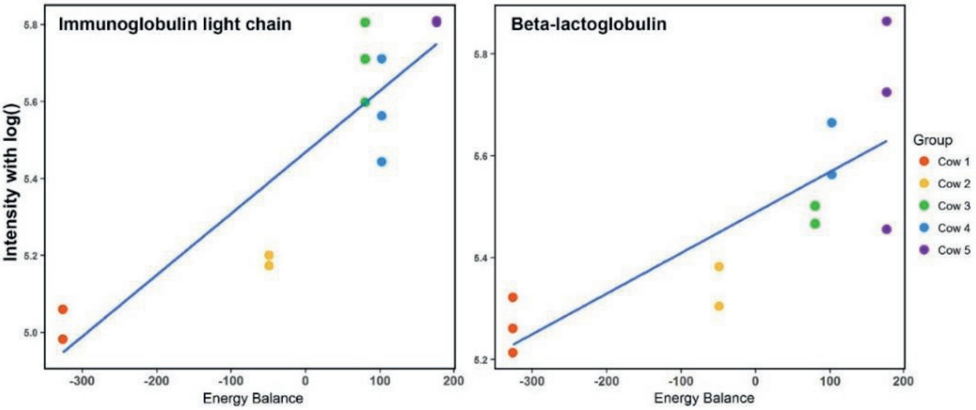


Fig. 5.1. The scatter plots of energy balance of 5 individual cows with 2 milk proteins with altered methylation level. Each cow has triplication in proteomics measurement. The outlier is not presented in the plot.

We also observed differences in concentrations of proteins in the samples measured. Based on the Pearson correlation analysis of energy balance and LFQ intensities, energy balance was negatively correlated with 7 milk proteins legumain, alpha-2-HS-glycoprotein, vitamin D binding protein, glycoprotein 2, folate receptor alpha, peptidyl-prolyl cis-trans isomerase B, and apolipoprotein A-IV. It should be noted that also some keratins (which are likely contaminants) correlated with energy balance. Energy balance was positively correlated with 2 proteins related to the immune system, an immunoglobulin-like and immunoglobulin-kappa like protein (Fig. 5.2).

The 7 proteins shown to be upregulated in negative energy balance are corroborating our metabolic studies. Apolipoprotein A-IV has a role in VLDL secretion and catabolism. Plasma free fatty acids (FFA) that are mobilized from adipose tissue are negatively correlated with energy balance as discussed in chapter 4. Free fatty acids are incorporated into very-low density lipoproteins (VLDL) which can be transported to the mammary gland and used for milk fat synthesis. Energy balance was negatively correlated with folate receptor alpha in milk (r = -0.92). Folate receptor alpha is related to folate metabolism [246,247]. In Chapter 2, 3, and 4, we have indicated that there is an increased

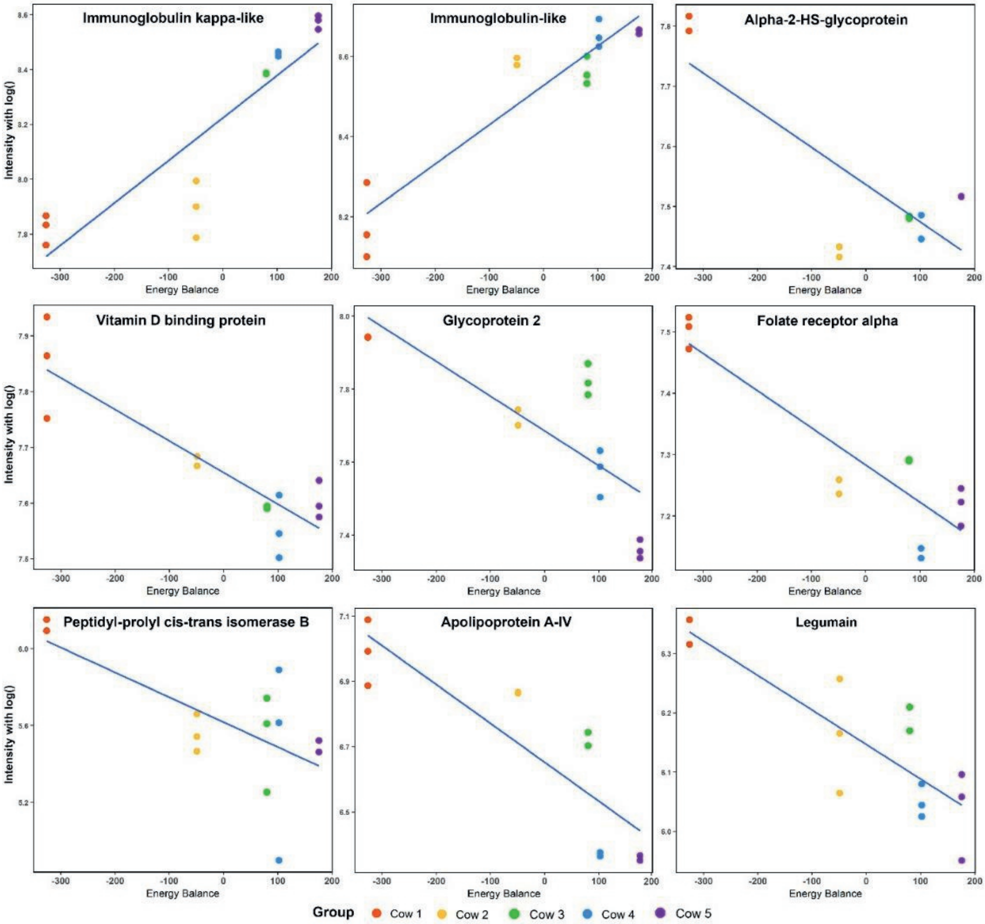


Fig. 5.2. The scatter plots of the correlation of the amount of 9 milk proteins with energy balance for 5 individual cows.

in DNA synthesis in the mammary gland of dairy cows in NEB due to an increase in cell proliferation. The negative correlation between energy balance and milk protein folate receptor alpha could further supports our findings on the protein level. Alpha-2-HS-glycoprotein, increased in cows in NEB, is reported to be promoting endocytosis and is important for regulation of bone mineralization (calcium). Legumain which is upregulated in cows in NEB is reported to have a cysteine-type endopeptidase activity and could play a role in cell division and regulation. Vitamin D binding protein is a multifunctional protein, involved in vitamin D binding and transport from plasma into the mammary gland cells. Glycoprotein 2 is an integral membrane protein, important in the innate immune response by binding pathogens such as enterobacteria. Peptidyl-prolyl cis-trans isomerase B is a protein involved in cis-trans isomerization of proline peptide bonds, a process important for protein folding of type I collagen. We proposed based on our metabolomics data (Chapter 2, 3 and 4) that proline and hydro-proline are upregulated in the mammary gland because of cell renewal with build-up of collagen for connective tissues. The upregulation of peptidyl-prolyl cis-trans isomerase B supports our hypothesis.

The down-regulation of two immunoglobulin-like proteins in NEB is surprising. In previous studies by Lu *et al.* (2013) it was shown that milk of cows in severe NEB showed higher concentrations of acute phase proteins related to the innate immune system [66]. Possibly the two immunoglobulin-like proteins are part of the adaptive immune system part of which might be downregulated in cows in NEB. Future studies will hopefully reveal in more detail the mechanistic relationship between the innate and adaptive immune system in cows in NEB.

5.5. Conclusions

This preliminary study is the first to show the correlation of energy balance of the methylation level of milk proteins and with the amount of milk proteins of individual dairy cows. Cows in negative energy balance had lower methylation status of two major abundant proteins in milk, immunoglobulin and β-lactoglobulin, which fits with the metabolomics data of cows in NEB, where we observed a decrease and a possible limitation in methyl donors for one carbon processes.

CHAPTER 6

Prediction of Metabolic Status of Dairy Cows in Early Lactation with On-farm Cow Data and Machine Learning Algorithms

Wei Xu[1,2], Ariette van Knegsel[1], Jacques Vervoort[2], Rupert Bruckmaier[4], Renny van Hoeij[1], Bas Kemp[1] & Edoardo Saccenti[3]

1. Adaptation Physiology Group, Department of Animal Sciences, Wageningen University & Research, the Netherlands

2. Laboratory of Biochemistry, Wageningen University & Research, the Netherlands

3. Laboratory of Systems and Synthetic Biology, Wageningen University & Research, the Netherlands

4. Veterinary Physiology, Vetsuisse Faculty, University of Bern. Switzerland

Journal of Dairy Science, 2019,102(11):10186-10201

6.1. Abstract

Metabolic status of dairy cows in early lactation can be evaluated through the concentrations of plasma *β*-hydroxybutyrate (BHB), free fatty acids (FFA), glucose, insulin, and insulin-like growth factor 1 (IGF-1). These plasma metabolites and metabolic hormones, however, are difficult to measure on farm. Instead, easily obtained on-farm cow data, such as milk production traits, have a potential to predict metabolic status. Here we aim, *i*) to investigate whether metabolic status of individual cows in early lactation could be clustered based on their plasma values, and *ii*) to evaluate machine learning algorithms to predict metabolic status using on-farm cow data. Through lactation week 1 to 7, plasma metabolites and metabolic hormones of 334 cows were measured weekly and used to cluster each cow into one of 3 clusters per week. The cluster with the greatest plasma BHB and FFA, and the lowest plasma glucose, insulin and IGF-1 was defined as a relatively poor metabolic status; the cluster with the lowest plasma BHB and FFA, and the greatest plasma glucose, insulin and IGF-1 was defined as a relative good metabolic status; the intermediate cluster was defined as a relative average metabolic status. Most dairy cows were attributed to an average or good metabolic status, limited number of cows were attributed to a poor metabolic status (10 to 50 cows per lactation week). On-farm cow data, including dry period length, parity, milk production traits, and body weight, were used to predict metabolic status with a good or an average metabolic status with 8 machine learning algorithms. Random Forest (error rate ranging from 12.4% to 22.6%) and Support Vector Machine (SVM, error rate ranging from 12.4% to 20.9%) are top 2 best performing algorithms to predict metabolic status using on-farm cow data. Random Forest had a higher sensitivity (range 67.8% to 82.9% during week 1 to 7) and negative predictive value (range: 89.5% to 93.8%), but lower specificity (range: 76.7% to 88.5%) and positive predictive value (range: 58.1% to 78.4%) than SVM. In Random Forest, milk yield, fat yield, protein percentage, and lactose yield had important roles in prediction, but their rank of importance differed across lactation weeks. In conclusion, dairy cows could be clustered for metabolic status, based on plasma metabolites and metabolic hormones. Moreover, on-farm cow data can predict cows in a good or an average metabolic status with best performance for Random Forest and SVM of all algorithms.

6.2. Introduction

In early lactation, high-yielding dairy cows suffer from body fat mobilization [28,248], and drastic metabolic changes [115,249]. Effective and accurate detection of metabolic status of dairy cows could benefit cow health and welfare by application of dietary and management strategies. Plasma metabolites and metabolic hormones have been used to indicate metabolic status of cows [20,250,251]. A poor metabolic status is related not only to increased plasma concentration of BHB and free fatty acids (FFA), and a decreased concentration of metabolic hormones, such as plasma insulin and insulin-like growth factor 1 (IGF-1) [252,253], but also to an increased incidence of metabolic disorders like ketosis and fatty liver [17,115,158]. Moreover, a poor metabolic status in early lactation, indicated by elevated plasma FFA or elevated BHB concentration, was associated with an increased risk for displaced abomasum, culling and reduced reproductive success, as reviewed by McArt *et al.* (2013) [254]. In addition, incidence of metabolic problems in early lactation has been related to an altered lactation curve with a lower peak milk yield and reduced 305-d milk production [255]. Blood sampling, however, is an invasive method, with limited applicability on the farm. A possible alternative is to predict metabolic status with on-farm cow data that are easily obtained, for example, body weight (BW), milk production traits, and parity. Milk production traits have been used to predict plasma BHB [256,257], and energy status at herd level of cows [23]. We hypothesize that on-farm cow data can predict metabolic status of dairy cows in early lactation.

Machine learning method uses statistical techniques to predict cow performance or disease events using large datasets [75,76]. On farm, available cow data usually include a series of correlated variables, for example, fat yield is correlated with milk yield [79]. Machine learning algorithms, however, can deal with complicated correlations caused by ever increasing number of variables [80], for example, Decision Tree has a good predictive performance by correlation-based variable selection [258], and Bayesian inference could mitigate the effect of correlations among a set of variables [259]. Machine learning algorithms, such as Naive Bayes, Random Forest, and Artificial Neural Networks (ANN), have been used to predict milk yield [81], breeding values [75], reproductive performance [76,83], and to identify mastitis [84,85] in dairy cows. Additionally, cluster analysis was used to define metabolic status in dairy cows in early lactation [174,260]. To our knowledge, however, the application of machine learning algorithms to predict metabolic status of dairy cows in early lactation using on-farm cow data has never been published. In this study, we aim, *i*) to define metabolic status of dairy cows through cluster analysis based on their plasma metabolites and metabolic hormones through lactation week 1 to 7; and *ii*) to predict metabolic status of those cows using on-farm cow data through machine learning algorithms.

6.3. Materials and Methods

6.3.1. Data and animals

Data was collected from two earlier studies with protocol number 2010026 (study I) [86,88] and 2014125 (study II) [126]. The experimental protocols were approved by the Institutional Animal Care and Use Committee of Wageningen University and were conducted at Dairy Campus research herd (WUR Livestock Research, Lelystad, the Netherlands). In study I, 168 Holstein-Friesian cows were assigned randomly to 1 of 3 groups with 0-d, 30-d or 60-d dry period length (DPL), 64 cows with a 0-d DPL (parity 2: 10 cows; parity ≥ 3: 54 cows); 69 cows with a 30-d DPL (parity 2: 11 cows; parity ≥ 3: 58 cows); 74 cows with a 60-d DPL (parity 2: 10 cows; parity ≥ 3: 64 cows). After one lactation within the experiment, 130 cows were monitored for a second lactation within study I [88]. In study II, 127 Holstein-Friesian cows were assigned randomly to 1 of 2 groups with either a 0-d or 30-d DPL: 84 cows with a 0-d DPL (parity 2: 40 cows; parity ≥ 3: 44 cows); 43 cows with a 30-d DPL (parity 2: 19 cows; parity ≥ 3: 24 cows). During all lactation stages in both study I and II, cows were housed in a free stall with slatted floor and cubicles, and cows were milked twice daily. Before calving, dry cows were fed a dry cow diet that consisted of grass silage, corn silage, wheat straw, rapeseed meal, urea, and vitamins and minerals (5.4 MJ NE_L/kg DM), whereas lactating cows were fed a lactation diet supporting 25 kg of milk. All cows were fed 1 kg/day of glucogenic or lipogenic concentrate from 10 days before the expected calving date. Postpartum, the experimental concentrate supply was increased in increments of 0.5 kg/day until a level of 8.5 kg/day was reached. Details on diet composition were presented earlier [86,88,126].

6.3.2. Measurements

Body weight and milk production traits. Body weight and milk yield were recorded daily and averaged per week. Milk fat, protein, and lactose percentage, and SCC were analysed per cow per week [(ISO9622, 2013), Qlip, Zutphen, the Netherlands]. Fat- and protein-corrected milk (FPCM) was calculated as:

$$FPCM\ (kg) = [0.337 + 0.116 \times fat\ (\%) + 0.06 \times protein\ (\%)] \times milk\ yield\ (kg)$$ [167]

Blood collection and analysis. During week 1 to 7 postpartum, 10 mL of blood was sampled every Thursday morning from the coccygeal vein into evacuated EDTA tubes (Vacuette, Greiner BioOne, Kremsmunster, Austria). Blood samples were kept on ice before centrifugation for plasma isolation (3,000 × *g* for 15 min, 4 °C). Plasma samples were stored at -20 °C before analysis. Plasma analysis has been previously described [261,262]. In short, the concentration of BHB was measured with kit no. RB1007 (Randox Laboratories, Ibach, Switzerland), FFA with kit no. 994-75409 (Wako Chemicals, Neuss, Germany), glucose with kit no. 61269 (BioMerieux, Marcy l'Etoile, France) [262], insulin with kit no. PI-12K (EMD Millipore Corporation, Billerica, MA, USA), and IGF-1 with kit no. A15729 (Beckman Coulter, Fullerton, CA, USA) [261].

6.3.3. Data transformation

Plasma glucose displayed a normal distribution in all lactation weeks. Raw data of FFA, BHB, insulin, and IGF-1 were first log transformed to correct for skewness.

6.3.4. Principal component analysis and correlation analysis

Principal component analysis (PCA) was performed to transform the data of 5 plasma metabolites and hormones into several orthogonal principal components (PC, [263]) for visualization and interpretation purposes. Data was centered and scaled to unit variance before analysis. The number of PC to retain was determined using the "elbow rule" [264]. The bi-plot (by 2-dimension) was used to investigate the patterns of variation of data and visualize the clustering with respect to different metabolic statuses. The correlation coefficient (r) among 5 plasma metabolites and hormones was determined by Pearson correlation.

6.3.5. Cluster analysis

In the current study, we assumed that the different metabolic status of dairy cows could be grouped through cluster analysis per lactation week. From study I and II, a total of 220, 315, 311, 310, 301, 299, and 298 cows with complete records for both blood and milk were used through lactation week 1 to 7 (Table 6.1). To assign dairy cows to metabolic status in each lactation week, *K*-means cluster analysis was applied on weekly plasma BHB, FFA, glucose, insulin, and IGF-1. Briefly, prerequisite cluster number *k* was determined by “Average Silhouette” method [265]. Then a set of *N* samples $X_{(1, 2.., i, i+1, ..., n)}$ were clustered into different $C_{j\,(1, 2...., k)}$, mean (μ_k) of all samples in each cluster is called the cluster “centroid”. Based on an iteration (maximum number is 10) that puts one sample into a cluster each time, *K*-means algorithm aims to choose the centroid with minimized within-cluster sum of squared Euclidean distance among each observation. Average values for plasma BHB, FFA, glucose, insulin and IGF-1 were calculated per cluster per lactation week and over all 7 lactation weeks.

6.3.6. Definition of metabolic clusters

Labelling of clusters for poor, average or good metabolic status was based on average value within a cluster for concentrations of plasma BHB, FFA, glucose, insulin and IGF-1 relative to the other clusters within the same lactation week. The cluster with the greatest plasma BHB and FFA, and the lowest plasma glucose, insulin and IGF-1 was defined as a relatively poor metabolic status; the cluster with the lowest plasma BHB and FFA, and the greatest plasma glucose, insulin and IGF-1 was defined as a relatively good metabolic status; the intermediate cluster was defined as a relative average metabolic status.

6.3.7. Machine learning algorithms

Machine learning algorithms were evaluated to predict cows in a poor, or an average, or a good metabolic status. Preliminary analysis indicated a limited number of cows in the poor metabolic status group; therefore, we build further classifiers to predict cows in either a good or an average metabolic status only. In each lactation week, models were trained by 8 machine learning algorithms using on-farm cow data with 13 features, including BW, DPL, parity, milk yield, milk yield SD, fat percentage, protein percentage, lactose percentage, FPCM, SCC, fat yield, protein yield, and lactose yield. Features were presented as an average per week, except milk yield which was included both as an average per week and as SD of milk yield per week. In the current study, cow in one lactation was regarded as independent to the same cow in another lactation. To evaluate several algorithms on a specific task is a common practice in machine learning, because performance of algorithm may depend on features, sample size, structure, and other characteristics of the dataset. In this study, 8 widely used machine learning algorithms were evaluated, including Decision Tree, Naive Bayes, Bayesian Network, Support Vector Machine, Artificial Neural Networks, Bootstrap Aggregation, Random Forest, and K-Nearest Neighbour. All methods were used with default parameters.

Table 6.1. The number of dairy cows with complete blood and milk observations in Study I and Study II through lactation week 1 to 7.

Study [1]	**Lactation week**						
	1	2	3	4	5	6	7
Study I							
First lactation	58	89	92	92	89	89	90
Second lactation	71	104	99	99	91	88	86
Study II	91	122	120	119	121	122	122
Total	220	315	311	310	301	299	298

[1] The experimental design, dry period length, and diet in Study I first lactation was described by van Knegsel *et al.* (2014) [86], Study I second lactation by Chen *et al.* (2016) [88], and in Study II by van Hoeij *et al.* (2017) [126].

6.3.8. Model evaluation

In each lactation week, models trained with eight machine learning algorithms were evaluated through 10-fold crossed validation. Briefly, in 10-fold crossed validation, 10% of the samples in whole dataset were kept as testing dataset, and the rest of the samples (90%) were used to train the model. Through lactation week 1 to 7, in total 29, 65, 64, 63, 56, 51, and 51 cows were in the dataset with 2 separate lactations. To avoid that the same cow is both in the test and the train dataset at the same time, with a different lactation number, a cow is only used once per loop of 10-fold cross validation. One loop ends until all samples were used exactly once in testing. For each machine learning algorithm, each loop of crossed validation repeated 500 time. Error rate is a priori value to evaluate the predictive performance of machine learning algorithms in each lactation week. To predict cows with 2 metabolic status, the error rate is defined as the ratio of false positive-to-false negative; to predict cows with 3 metabolic status:

$$Error\ rate = 1 - (sum\ of\ corrected\ prediction\ /\ sample\ size) * 100\%$$

Moreover, the sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of machine learning algorithms in each lactation week were also presented [266]. Based on preliminary analysis, Random Forest performed well not only in mode accuracy, but also in sensitivity and NPV. The importance of each feature in Random Forest was investigated by mean decreased Gini [267]. Briefly, importance of feature X_m to predict Y is evaluated by adding up weighted impurity decreases $p(t)\ \Delta i\ (s_t, t)$ for all nodes t where X_m is used, averaged over all N_T trees in Random Forest:

$$Imp\ (X_m) = \frac{1}{N_T} \sum_{T} \sum_{t \in T: v(s_t) = X_m} p(t)\ \Delta i(s_t, t) \qquad (Equation\ 2)$$

and where X_m is the features used in Random Forest, $p(t)$ is the proportion N_t/N of samples reaching t, and $v(s_t)$ is the variable used in split s_t [268]. A lower Gini (i.e. higher decreased Gini) means that a feature plays a relatively more important role than others to discriminate samples into defined classes.

6.3.9. Data processing and software

Multiple comparison analysis was applied with Tukey Honest Significant Differences test. Data pre-processing and machine learning algorithms were programmed in the R environment (version 3.3.3), with function "kmeans()" for cluster analysis, function "prcomp()" for principal component analysis, function "cor.test()" for Pearson analysis, package "rpart" for Decision Tree, "neuralnet" for Bayesian Network, "e1071" for Naïve Bayes and Support Vector Machine, "adabag" for Bootstrap Aggregation, and "randomforest" for Random Forest. Figures were plotted in Python (version 3.6), using modules "pandas", "matplotlib", and "Seaborn".

6.4. Results

6.4.1. Principal component analysis and correlation analysis

All plasma metabolites and metabolic hormones were correlated (Pearson) with a r (absolute value) less than 0.61 ($P < 0.05$) or insignificant ($P > 0.05$) (Table 6.2). Fig. 6.1 shows the bi-plot (limited to the first two components, for sake of simplicity) from the PCA and presents the explained variation in 1st principal component (PC, range 47.5% to 62.6% during week 1 to 7) and 2nd PC (range: 13.5% to 18.2%). Inspection of the PCA loadings shows the relative contribution of 5 plasma metabolites and hormones to cluster metabolic status of dairy cows (bi-plot of PCA, Fig. 6.1). Raw data of loading is presented in *Appendix* Table 6.1. The loading pattern of plasma metabolites and hormones is the same across lactation week 1 to 7.

6.4.2. Definition of clusters for metabolic status

Optimal number of clusters suggested by Average Silhouette was “3” through lactation week 1 to 7. In each lactation week, the silhouette widths per k numbers in Cluster Analysis were presented in *Appendix* Fig. 6.1. Based on the concentration of plasma metabolites and metabolic hormones, cows in the first, second, and third cluster were defined as having a good, average, and a poor metabolic status, respectively. Cluster 3 defined as poor metabolic status had the greatest plasma BHB and FFA, but the lowest glucose, insulin and IGF-1 ($P < 0.05$, Table 6.3); conversely, cluster 1 defined as good metabolic status had the lowest plasma BHB and FFA, but the greatest plasma glucose, insulin and IGF-1 ($P < 0.05$). For all lactation weeks, cluster 2 had a lower plasma BHB or lower plasma FFA than cluster 3 ($P < 0.05$), but also a lower plasma insulin, IGF-1 and glucose than cluster 1 ($P < 0.05$), and was therefore defined as the cluster with cows with an average metabolic status.

6.4.3. Predictive performance of machine learning algorithms

To predict cows in either a good or an average metabolic status, SVM had the lowest error rate in lactation week 3 to 7 of all algorithms. Random Forest had the lowest error rate in lactation week 1 and 2, and slightly higher error rate through lactation week 3 to 7 (Table 6.3). The sensitivity and NPV of all algorithms are relative lower in lactation week 1 than other weeks. Random Forest had relative higher sensitivity (range from 67.8% to 82.9% during week 1 to 7) and NPV (range: 89.5% to 93.8%), and SVM had a relative higher specificity (range: 80.9% to 93.7%) and PPV (range: 78.8% to 86.1%) (Fig. 6.2). After 500 repetitions in Random Forest, error rate to predict cows in a good metabolic status was higher than error rate to predict cows in an average metabolic status (Fig. 6.1). In two-dimensional PCA plots, major incorrectly predicted dots (big size dots) are mainly presented in border area between good and average metabolic status groups (Fig. 6.1).

Machine learning algorithms to predict cows in a poor, or an average, or a good metabolic status were evaluated (*Appendix* Table 6.2). The general performance of 8 algorithms decreased when 3 metabolic statuses are included in the prediction, compared with prediction of 2 metabolic statuses, error rate ranged from 21.36% to 48.7%. Random Forest and SVM are still best performed algorithm through all lactation weeks.

Table 6.2. Pearson coefficients ($P < 0.05$) among the concentration of plasma free fatty acids (FFA), BHB, glucose, insulin, and IGF-1 of dairy cows through lactation week 1 to 7. (n.s., not significant)

	week 1				week 2			
	FFA	BHB	Glucose	Insulin	FFA	BHB	Glucose	Insulin
BHB	0.52				0.52			
Glucose	-0.18	-0.36			-0.49	-0.53		
Insulin	-0.47	-0.20	0.32		-0.38	-0.24	0.48	
IGF-1	-0.55	-0.29	0.34	0.58	-0.51	-0.31	0.60	0.50
	week 3				week 4			
	FFA	BHB	Glucose	Insulin	FFA	BHB	Glucose	Insulin
BHB	0.41				0.39			
Glucose	-0.48	-0.57			-0.45	-0.53		
Insulin	-0.44	-0.20	0.45		-0.38	-0.24	0.40	
IGF-1	-0.50	-0.30	0.61	0.45	-0.52	-0.16	0.57	0.32
	week 5				week 6			
	FFA	BHB	Glucose	Insulin	FFA	BHB	Glucose	Insulin
BHB	0.43				0.39			
Glucose	-0.43	-0.52			-0.37	-0.48		
Insulin	-0.33	-0.17	0.40		-0.20	n.s.	0.32	
IGF-1	-0.49	-0.25	0.57	0.37	-0.42	-0.22	0.53	0.25
	week 7							
	FFA	BHB	Glucose	Insulin				
BHB	0.32							
Glucose	-0.38	-0.43						
Insulin	-0.29	-0.17	0.34					
IGF-1	-0.46	-0.18	0.57	0.31				

6.4.4. Feature contribution

Random Forest had the relative lower error rate, but also had highest sensitivity and NPV to predict dairy cows with a good metabolic status (Table 6.4 and Fig. 6.2). Contribution of features to predict metabolic status in Random Forest is shown in Fig. 6.3. Through lactation week 1 to 3, milk yield, fat yield, and protein percentage were more important than other features, in week 4 and 5, protein percentage and lactose yield were top important features to predict metabolic status; in week 6 and 7, milk yield, protein percentage, and lactose yield are more important than other features, while fat related features were not as important as in lactation week 1 to 3.

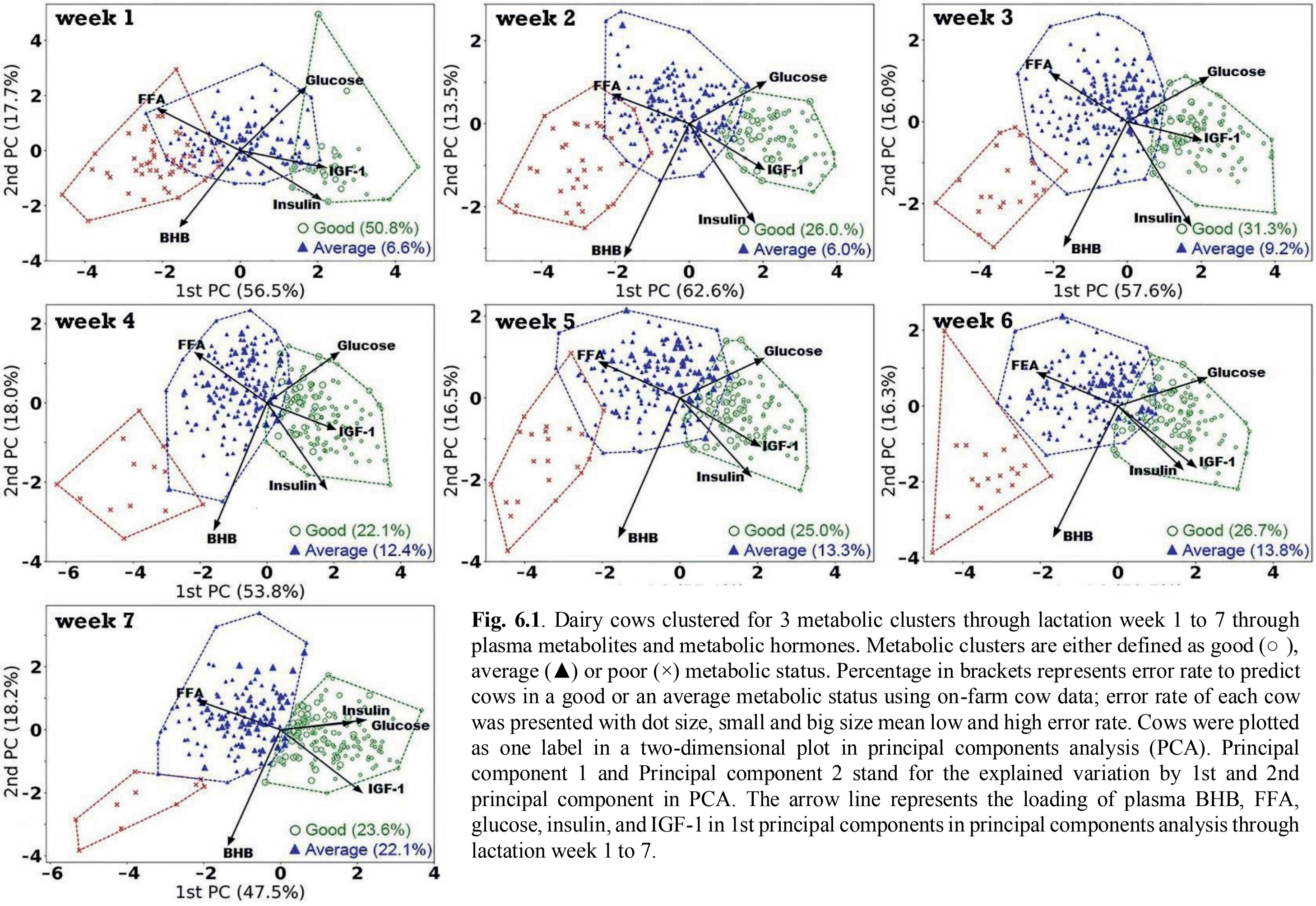


Fig. 6.1. Dairy cows clustered for 3 metabolic clusters through lactation week 1 to 7 through plasma metabolites and metabolic hormones. Metabolic clusters are either defined as good (○), average (▲) or poor (×) metabolic status. Percentage in brackets represents error rate to predict cows in a good or an average metabolic status using on-farm cow data; error rate of each cow was presented with dot size, small and big size mean low and high error rate. Cows were plotted as one label in a two-dimensional plot in principal components analysis (PCA). Principal component 1 and Principal component 2 stand for the explained variation by 1st and 2nd principal component in PCA. The arrow line represents the loading of plasma BHB, FFA, glucose, insulin, and IGF-1 in 1st principal components in principal components analysis through lactation week 1 to 7.

Table 6.3. Plasma metabolite and metabolic hormone concentrations for dairy cows (number, N) in 3 metabolic clusters (C) in lactation week (wk) 1 to 7 postpartum. Cluster analysis of metabolic status (MS) was based on the concentration of plasma free fatty acids (FFA), BHB, glucose, insulin, and IGF-1. Values represent means (SD).

wk	C	N	BHB (mmol/L)	FFA (mmol/L)	Glucose (mmol/L)	Insulin (µU/mL)	IGF-1 (ng/mL)	MS
1	1	35	0.60 (0.17)^{a}	0.12 (0.10)^{a}	4.63 (1.27)^{a}	31.81 (17.88)^{a}	162.9 (55.0)^{a}	Good
	2	135	0.58 (0.13)^{a}	0.38 (0.22)^{b}	3.75 (0.67)^{b}	10.34 (4.81)^{b}	71.7 (36.6)^{b}	Average
	3	50	1.10 (0.36)^{b}	0.87 (0.32)^{c}	3.25 (0.59)^{c}	6.37 (3.44)^{c}	33.6 (20.9)^{c}	Poor
		Average	*0.70 (0.30)*	*0.45 (0.33)*	*3.78 (0.86)*	*12.87 (11.72)*	*77.6 (54.9)*	
2	1	80	0.56 (0.13)^{a}	0.14 (0.11)^{a}	3.90 (0.29)^{a}	18.35 (7.90)^{a}	154.3 (69.1)^{a}	Good
	2	199	0.75 (0.26)^{b}	0.43 (0.21)^{b}	3.22 (0.36)^{b}	7.96 (3.75)^{b}	57.1 (28.9)^{b}	Average
	3	36	2.55 (1.25)^{c}	0.97 (0.43)^{c}	2.55 (0.43)^{c}	5.55 (2.52)^{c}	32.2 (16.4)^{c}	Poor
		Average	*0.91 (0.76)*	*0.42 (0.32)*	*3.32 (0.53)*	*10.32 (6.92)*	*78.9 (61.3)*	
3	1	95	0.60 (0.14)^{a}	0.13 (0.09)^{a}	3.86 (0.29)^{a}	18.72 (9.19)^{a}	138.9 (49.5)^{a}	Good
	2	195	0.80 (0.32)^{b}	0.40 (0.19)^{b}	3.24 (0.36)^{b}	8.86 (4.28)^{b}	69.7 (26.6)^{b}	Average
	3	21	2.95 (0.91)^{c}	0.64 (0.24)^{c}	2.48 (0.44)^{c}	7.07 (3.39)^{b}	47.6 (19.5)^{c}	Poor
		Average	*0.89 (0.66)*	*0.33 (0.22)*	*3.38 (0.51)*	*11.75 (7.71)*	*89.4 (48.2)*	
4	1	129	0.64 (0.19)^{a}	0.12 (0.08)^{a}	3.88 (0.30)^{a}	19.2 (10.9)^{a}	131.9 (41.9)^{a}	Good
	2	168	0.79 (0.37)^{b}	0.39 (0.20)^{b}	3.33 (0.36)^{b}	9.70 (4.06)^{b}	73.5 (25.8)^{b}	Average
	3	13	3.79 (1.25)^{c}	0.67 (0.24)^{c}	2.47 (0.55)^{c}	8.01 (6.58)^{b}	52.3 (29.2)^{b}	Poor
		Average	*0.85 (0.73)*	*0.28 (0.22)*	*3.52 (0.49)*	*13.58 (9.08)*	*96.9 (44.9)*	
5	1	110	0.63 (0.22)^{a}	0.11 (0.08)^{a}	3.99 (0.32)^{a}	22.03 (12.63)^{a}	140.4 (41.5)^{a}	Good
	2	169	0.71 (0.30)^{a}	0.30 (0.16)^{b}	3.48 (0.34)^{b}	10.37 (4.23)^{b}	79.5 (26.4)^{b}	Average
	3	22	2.97 (1.54)^{b}	0.63 (0.33)^{c}	2.72 (0.35)^{c}	8.11 (3.83)^{c}	53.1 (16.4)^{b}	Poor
		Average	*0.85 (0.77)*	*0.26 (0.21)*	*3.61 (0.48)*	*14.47 (10.12)*	*99.8 (45.1)*	
6	1	117	0.60 (0.19)^{a}	0.11 (0.08)^{a}	4.03 (0.28)^{a}	20.69 (14.01)^{a}	143.4 (40.1)^{a}	Good
	3	163	0.67 (0.25)^{a}	0.27 (0.16)^{b}	3.55 (0.31)^{b}	11.64 (5.78)^{b}	76.6 (23.9)^{b}	Average
	3	19	2.76 (1.27)^{b}	0.51 (0.37)^{c}	2.97 (0.31)^{c}	8.96 (2.44)^{b}	62.0 (22.5)^{b}	Poor
		Average	*0.78 (0.64)*	*0.22 (0.19)*	*3.70 (0.42)*	*15.01 (10.78)*	*101.8 (45.7)*	
7	1	144	0.61 (0.20)^{a}	0.09 (0.06)^{a}	3.99 (0.28)^{a}	20.2 (12.92)^{a}	132.8 (37.0)^{a}	Good
	2	144	0.69 (0.29)^{a}	0.25 (0.14)^{b}	3.55 (0.31)^{b}	12.02 (5.97)^{b}	74.3 (23.6)^{b}	Average
	3	10	3.08 (1.25)^{b}	0.37 (0.11)^{c}	2.97 (0.58)^{c}	8.30 (2.97)^{b}	68.7 (15.5)^{b}	Poor
		Average	*0.73 (0.55)*	*0.18 (0.14)*	*3.74 (0.40)*	*15.85 (10.77)*	*102.4 (42.5)*	
Average wk 1 till 7			*0.61 (0.18)*^{a}	*0.11 (0.08)*^{a}	*3.98 (0.43)*^{a}	*20.55 (12.38)*^{a}	*140.3 (46.9)*^{a}	Good
			0.72 (0.29)^{b}	*0.35 (0.20)*^{b}	*3.42 (0.43)*^{b}	*9.99 (4.88)*^{b}	*71.3 (28.3)*^{b}	Average
			2.38 (1.27)^{c}	*0.75 (0.37)*^{c}	*2.83 (0.57)*^{c}	*7.03 (3.66)*^{c}	*44.2 (23.2)*^{c}	Poor

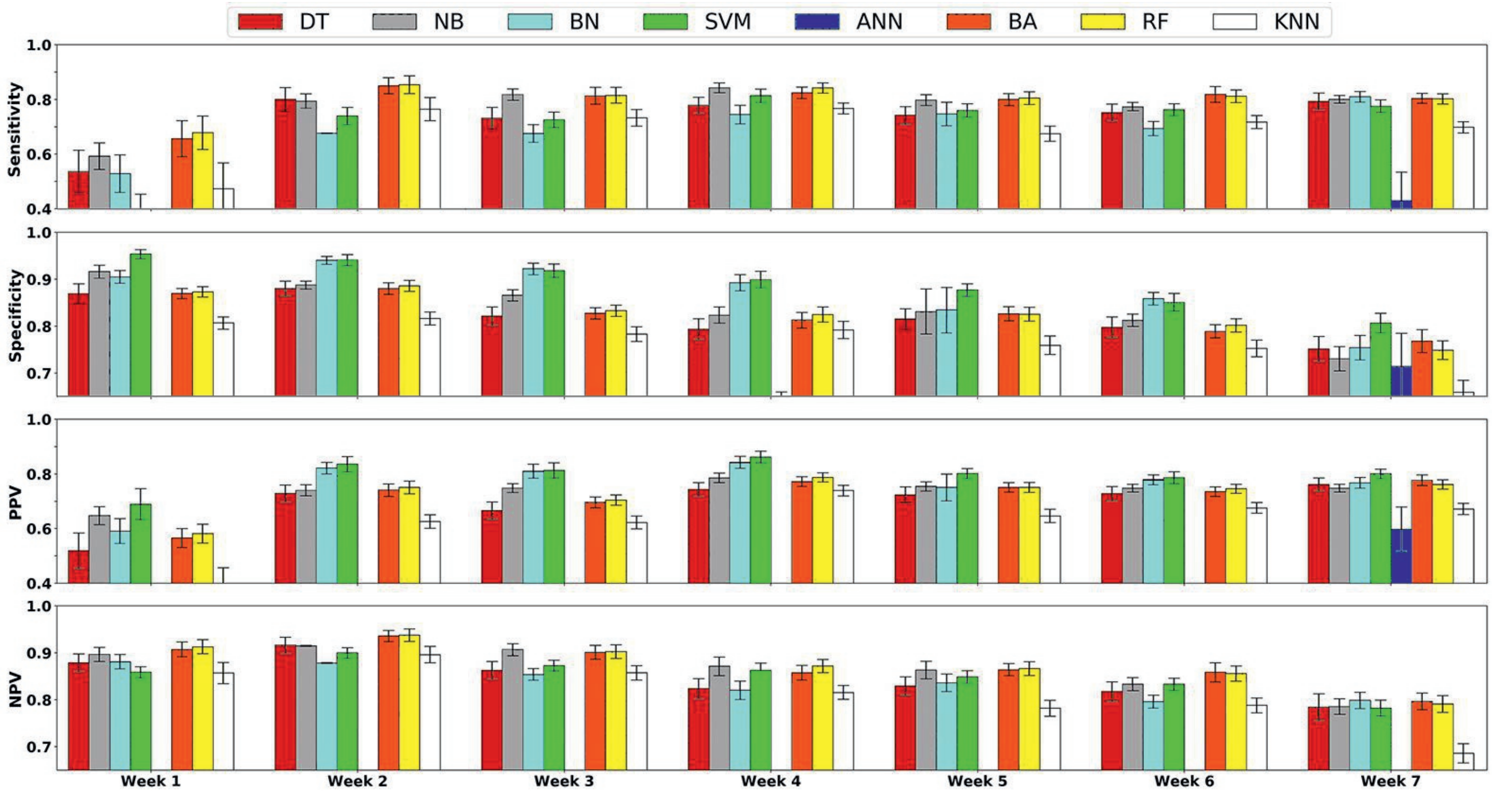


Fig. 6.2. The sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of machine learning algorithms [1] to predict cows in a good or an average metabolic status using on-farm cow data through lactation week 1 to 7.

Abbreviations: ANN, Artificial Neural Networks; BA, Bootstrap Aggregation; BN, Bayesian Network; DT, Decision Tree; KNN, K-Nearest Neighbour; NB, Naive Bayes; SVM, Support Vector Machine; RF, Random Forest.

[1] The sensitivity and PPV of ANN are not presented in Fig. 6.2 due to averages lower than 0.40 in all lactation weeks; the specificity and NPV of ANN are not presented in Fig. 6.2 due to averages lower than 0.65 in all lactation weeks.

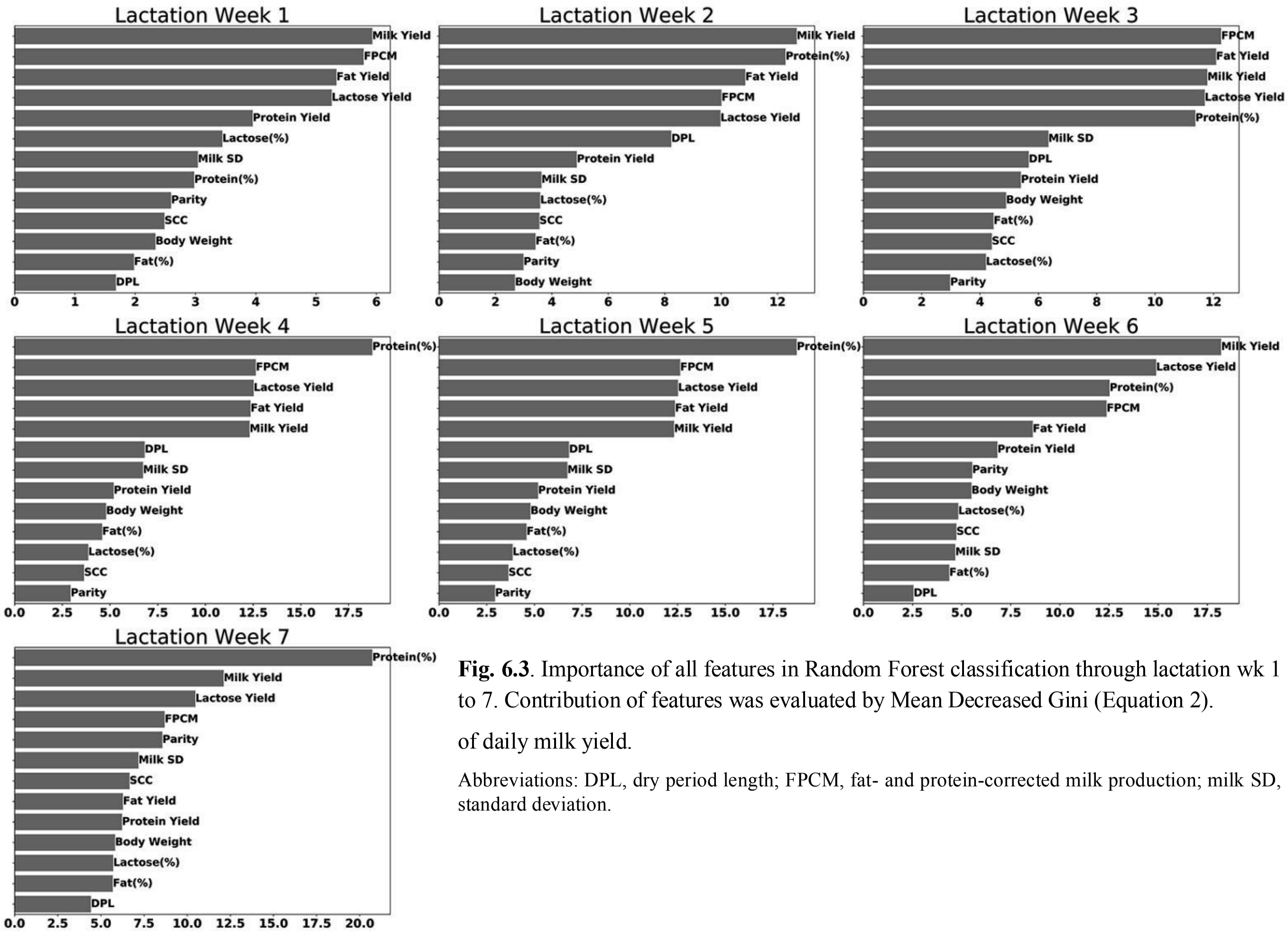


Fig. 6.3. Importance of all features in Random Forest classification through lactation wk 1 to 7. Contribution of features was evaluated by Mean Decreased Gini (Equation 2).

of daily milk yield.

Abbreviations: DPL, dry period length; FPCM, fat- and protein-corrected milk production; milk SD, standard deviation.

6.5. Discussion

6.5.1. Metabolic clusters

In the current study, we first aimed to define metabolic status of dairy cows through cluster analysis based on their plasma metabolites and metabolic hormones through lactation week 1 to 7. It was hypothesized that altered plasma metabolites and metabolic hormones between different clusters could indicate the difference in metabolic status. Moreover, the correlation among these plasma variables were in line with previous studies, for example, greater plasma glucose concentration was negatively correlated with plasma BHB and FFA concentration [269,270]. In the current study, however, majority of the correlations between plasma metabolites and metabolic hormones had a trend to get weak from lactation week 1 to 7. For example, both the correlations beween plasma FFA and insulin and plasma IGF-1 and insulin got smaller from week 1 till 7. Cluster analysis can be affected by variables with high correlation [271], which is usually defined as a correlation of above 0.8 or 0.9 [272]. Correlations among variables in the current study were maximal 0.61. An alternative could be to perform cluster analysis on selected principal components, which are orthogonal and thus uncorrelated [260,273]. The sum explained variation by first and second PC accounted for 65.7% to 76.1% over lactation weeks in the current study (Fig. 6.1). Therefore, when using selected PCs (for example, 2 PCs) from PCA in cluster analysis, 2 PCs could only account for 65.7% to 76.1% explained variation of 5 plasma metabolites and metabolic hormones.

In the current study, metabolic status was a composite trait of plasma BHB, FFA, glucose, insulin and IGF-1. Cows in a poor metabolic status had on average a greater BHB concentration than the earlier suggested threshold for subclinical ketosis (>1.2 mmol/L [274]), or on average a greater plasma FFA than the suggested threshold for subclinical ketosis (>0.6 mmol/L) [17], except in lactation week 7. Recently, Tremblay *et al.* (2018) identified plasma values for poor metabolic adaptation syndrome (PMAS) in dairy cows in early lactation [260]. Results of this study are in line with the current study for cows with a good metabolic status, but not completely for cows with a poor metabolic status. In the current study, cows with a good metabolic status had over lactation weeks on average a lower plasma FFA concentration than the earlier suggested threshold for low PMAS (<0.39 mmol/L) [260]. Dairy cows with a poor metabolic status, however, had in week 1 and 2 on average a greater plasma FFA concentration than the earlier suggested threshold for high PMAS (≥0.7 mmol/L) [260]. Difference between these 2 studies might be related to the fact that PMAS clusters were based on plasma FFA, BHB and milk fat-to-protein ratio [260], while in the current study metabolic status was a composite trait based on plasma FFA, BHB, glucose, insulin and IGF-1. For all lactation weeks, cows with a relative good metabolic status had plasma IGF-1 concentration >100 ng/mL [275], and greater plasma glucose and insulin concentration than the other 2 clusters. Moreover, in our study, cluster analysis was done per lactation week, which avoids a possible confounding effect of lactation stage with metabolic status of dairy cows. Implication of this approach is that metabolic health classifications are relative to the other 2 clusters within a lactation week. This could imply that what is defined as a poor metabolic status in week 2 is different from what is defined as a poor metabolic status in week 7. Comparing, however, cluster 1 and 3 over lactation weeks, shows that over lactation weeks the differences among clusters are marginal, indicating that the cows defined as having a poor metabolic status have a similar metabolic profile over the first 7 lactation weeks in the current study.

In bi-plot of PCA, the opposing loadings of 5 plasma metabolites and metabolic hormones could indicate how these values contribute the separation of dairy cows with a relative good, average, or poor metabolic status. The negative contribution of plasma BHB and FFA to 1st principal component means that high BHB and FFA are related to a poor metabolic status in cluster analysis. Concentration of FFA is an indicator of fat mobilization during poor metabolic status, while plasma BHB is an indicator of fat metabolism and incomplete oxidation of fat [115]. In our study, cows in poor metabolic status had the lowest concentration of glucose, insulin and IGF-1 among 3 clusters. Glucose in early lactation is not only essential for lactose production, but also to metabolize fatty acids and make energy available to the body in the form of ATP [156]. Low insulin reduces glucose uptake by insulin-responsive peripheral tissues (adipose and muscle) and facilitate greater uptake of glucose by the mammary gland [276]. Low concentration of IGF-1 and insulin is related to the NEB status of cows [277]. Cows with the greatest concentration of insulin and IGF-1 were defined as having a good metabolic status, which fits with the observation that great concentrations of insulin and IGF-1 were found in cows with positive energy balance [277,278]. Associated with increasing plasma glucose and insulin availability, increased IGF-1 can be hypothesized to have anabolic effects on carbohydrate metabolism [279,280]. Cows in a good metabolic status had greater IGF-1 (Table 6.2), lower body weight and lower average parity (*Appendix* Table 6.3). In mammals, IGF-1 promotes growth and ageing [281,282], which could indicate that cows in a good metabolic status have a more anabolic status. Therefore, cows in good metabolic status have low body weight, milk yield, protein yield, and fat yield (*Appendix* Table 6.3). The low average BW and low average parity in this cluster suggest there might be a greater proportion of young cows in this cluster, which still have some priority for growth, which might explain the greater plasma IGF-1 concentration for these cows.

In the current study, machine learning algorithms were used to predict cows in a good or average metabolic status, but not for cows in a poor metabolic status. This approach was chosen due to a limited number of animals in the cluster with a poor metabolic status in order to improve the predictive performance of the algorithms. For cow health management, however, prediction of cows in a poor metabolic status is very informative. Therefore, evaluation of machine learning algorithms to predict cows in a poor, average or good metabolic status is added in *Appendix* Table 6.2. Identification of metabolic status could facilitate cow health management by fine-tuning of dietary and management strategies for individual dairy cows in a precision farming system, as reviewed by Rutten *et al*. (2013) [283]. On the one side, cows in a poor metabolic status could be supplied with a diet with greater energy content, or attributed to a 30-d or 0-d DPL to improve metabolic status in the next lactation. On the other side, cows in a good metabolic status might deserve extra attention to prevent metabolic diseases in the next lactation caused by body fat accumulation in the current lactation [115]. First, limiting energy supply could result in reduced body fat storage, and lower BCS [284]. Second, dairy cows with a good metabolic status are candidate for a 30-d or 60-d DPL to reduce energy balance and maximized milk yield in the next lactation [88].

6.5.2. Predict metabolic status with machine learning algorithms

In this study, *K*-means was an unsupervised learning method in cluster analysis, whereas KNN was a supervised learning method to predict metabolic status. Although both algorithms used “Euclidean Distance” as kernel function, KNN could limitedly predict metabolic status with on-farm cow data. It is difficult to consider a specific linear function with conventional statistical methods, due to its limitations when using correlated variables. In this context, analysing data through machine learning algorithms seems promising to manage large datasets with on-farm cow data with a lot of correlations among features [285].

The superior performance of Random Forest and Bootstrap Aggregation is in line with study that predicts insemination outcomes in cows [83], which could be explained by the power of ensemble methods to generate high-performance classifiers by collecting individual trained classifiers. Although Random Forest performed better than other algorithms, even to predict cows with 3 metabolic statuses in our study (*Appendix* Table 6.2), it does not mean Random Forest will always be the best algorithm in practice. The performance of all algorithms, for example Random Forest, depends on used features and parameters. A multi-parallel comparison among algorithms, therefore, is necessary to evaluate algorithms in a specific case.

Other machine learning algorithms had relatively higher error rate than Random Forest and Bootstrap Aggregation. Among all algorithms, ANN is frequently used, for example, to predict milk yield [81,286], and to classify mastitis cases [84,85]. In our study, however, the error rate of ANN (21.3% to 39.2%, Table 6.3) in prediction of metabolic status is high, compared with other studied algorithms. Reasons for difference in fit of ANN among studies are unknown.

All algorithms had a better prediction of metabolic status of cows in lactation week 2 and 3, compared with prediction in week 1, and 4 to 7, which could be explained by the dramatic changes of metabolic status at the beginning of lactation [249]. Dramatic changes of metabolic status in cows with low adaptive capacity in early lactation could benefit the prediction of metabolic status by machine learning algorithm in the current study. In addition, major incorrectly predicted cases in Random Forest were presented in border area between good and an average metabolic status of cows in the PCA plot (Fig. 6.1). This result indicates that on-farm cow data are limited to predict metabolic status of cows in the border area between clusters, which could be explained by some features that are not different between cows in a good or an average metabolic status, such as BW, fat percentage, lactose percentage, and FPCM (*Appendix* Table 6.3).

The performance of algorithms with different sensitivity, specificity, PPV, and NPV determines application in practice. The relative higher sensitivity and NPV value of Random Forest could be helpful to detect cows with surplus energy balance (also referred to a good metabolic status). In addition, it would be expected to further study the reason that machine learning algorithms perform different in sensitivity, specificity, PPV, and NPV, and improve the performance of applied algorithms in practice. For example, how the sensitivity of prediction of metabolic status could be improved by using dietary component data. Moreover, several other aspects can be hypothesized to facilitate further fine-tuning of machine learning algorithms to predict metabolic status. First, larger datasets and more comprehensive on-farm cow data or on-farm non-cow data. E.g. datasets including more cows in a poor metabolic status would allow more precise prediction of cows in a poor metabolic status; weather data could be used to predict energy intake of dairy cows [287], which could be related to metabolic status in early lactation. Second, in the current study, features were included as the weekly average or SD per week for milk yield. Also other on-farm cow data can be hypothesized to contribute to a model to estimate metabolic status like the minimum or maximum milk yield, feeding activity or lying time per day [174]. Third, models could be further optimized using not only data per lactation week, but also data of all lactation weeks at once. This would imply the random effect of individual cows in consecutive lactation weeks needs to be considered, which would request more complicated algorithms, which are, to our knowledge, not standard procedure. Fourth, machine learning algorithms can be used not only to predict metabolic status, like Tremblay *et al.* (2018) [260], or the current study, but specific diseases and disorders associated with metabolic status, like fatty liver or subclinical ketosis in dairy cows [288].

6.5.3. Important features

In the current study, Random Forest had the relative lower error rate, but also had highest sensitivity and NPV to predict 362 dairy cows with a relatively good metabolic status. In lactation week 1, 2 and 3, milk yield, fat yield, and protein percentage played important roles to predict cows in a good or an average metabolic status in Random Forest, which could be explained by relation between milk production and body fat mobilization. Increased fat yield in early lactation originates partly from mobilized body fat, this process is related to elevated plasma concentration of FFA and BHB [115]. Therefore, milk fat and fat-to-protein ratio were used to estimate energy balance of cows at herd level [18,140]. In the current study, in lactation week 4 and 5, protein percentage and lactose yield were top 2 important features in Random Forest. It can be speculated that milk protein and lactose are important to assign cows to a relative good or an average metabolic status. Cows in an average metabolic status had greater protein and lactose production than cows in a good and a poor metabolic status (*Appendix* Table 6.3). In lactation week 6 and 7, fat related features were not as important as milk yield, milk protein percentage, and lactose yield. In lactation week 6 and 7, cows in an average metabolic status had same fat yield as cows in a poor metabolic status (*Appendix* Table 6.3), which could be explained by the recovery to positive energy balance status after 41.5 DIM postpartum [27]. In expectation, fine tuning the model by adjusting the contribution of milk yield, fat yield, protein percentage, and lactose yield in each lactation week to the model could improve predictive performance in further studies.

6.6. Conclusions

Through lactation week 1 to 7, dairy cows were clustered for either a relatively poor, average, or good metabolic status, based on 5 plasma metabolites and metabolic hormones. On-farm cow data predicted cows in a good or an average metabolic status in early lactation using 8 machine learning algorithms. Error rate of those algorithms ranged from 12.4% to 49.8%. Of all algorithms, Random Forest and Support Vector Machine had stable and best performance to predict metabolic status of dairy cows. Milk yield, fat yield, protein percentage, and lactose yield had important roles in Random Forest, however, their rank of importance differed across lactation weeks.

CHAPTER 7

Prediction of Hyperketonemia in Dairy Cows in Early Lactation Using On-farm Cow Data and Net Energy Intake by Partial Least Square Discriminant Analysis

Wei Xu[1, 2], Edoardo Saccenti[3], Jacques Vervoort[2], Bas Kemp[1], Rupert Bruckmaier[4] & Ariette van Knegsel[1]

1. Adaptation Physiology Group, Department of Animal Sciences, Wageningen University & Research, the Netherlands

2. Laboratory of Biochemistry, Wageningen University & Research, the Netherlands

3. Laboratory of Systems and Synthetic Biology, Wageningen University & Research, the Netherlands

4. Veterinary Physiology, Vetsuisse Faculty, University of Bern. Switzerland

Submitted

7.1. Abstract

The objectives of this study were: *i*) to evaluate if hyperketonemia in dairy cows (defined as plasma BHB $\geq$ 1.0 mmol/L) can be predicted using on-farm cow data either in current or previous lactation week, and *ii*) to study if adding individual net energy intake (NEI) can improve the predictive ability of the model. Plasma BHB concentration, on-farm cow data, and NEI of 532 individual cows were available weekly through lactation week 1 to 5 postpartum. To predict hyperketonemia in dairy cows, models were first trained by partial least square discriminant analysis (PLS-DA), using on-farm cow data in same or previous lactation week. Second, NEI was included in models to evaluate the improvement of predictive ability of PLS-DA. Through 5,000 times 5-fold crossed validation, models were evaluated by accuracy (the ratio of the sum of true positive and true negative), sensitivity, specificity, positive predictive value, and negative predictive value. Through lactation week 1 to 5, the accuracy to predict hyperketonemia using data in same week was 75.9% to 83.8% (on-farm cow data only), 78.0% to 84.9% (model including NEI), and using data in previous week was 72.0% to 79.0% (on-farm cow data only), 72.8% to 80.7% (model including NEI). Improvement of the accuracy of the model due to including NEI ranged among lactation weeks from 0.6% to 3.7% when using data in the same lactation week and 0.6% to 1.6% when using data in the previous lactation week. In conclusion, trained models via PLS-DA have potential to predict hyperketonemia in dairy cows using data not only in the current lactation week, but also using data from the previous lactation week. Net energy intake can improve the accuracy of model, but only to a limited extent. Besides NEI, body weight, milk fat and protein content and milk fat and protein yield were important variables to predict hyperketonemia, but their rank of importance differed across lactation weeks.

7.2. Introduction

In early lactation, dairy cows typically have a negative energy balance (NEB) which has been related to metabolic disorders, like hyperketonemia [115,289]. Hyperketonemia is defined as an increased concentration of plasma BHB, the usually used threshold is plasma concentration of BHB $\geq$ 1.0 to $\geq$ 1.4 mmol/L in dairy cows [141,289]. The incidence of hyperketonemia is especially high (up to 45.7%) in the first weeks after calving [290]. Hyperketonemia is related to an increased risk of disorders in peripartum period, such as subclinical ketosis and ketosis [103,289], left-displaced abomasum [291], and decreased reproductive performance [141]. Reliable assessment of hyperketonemia could thus diagnose cows with metabolic disorder, but no clinical signs yet. Urine or milk ketone tests have been applied as cow-side tests to diagnose hyperketonemia in dairy cows on farms [96]. These cow-side tests, however, are not used to frequently screen all cows due to the costs and labour associated with these tests. Alternatively, plasma BHB has been predicted using milk fat-to-protein ratio [21], milk metabolites [64], and milk BHB and acetone [257]. Those methods estimated plasma BHB using data in same lactation week. A reliable and early prediction of hyperketonemia in cows could assist in cow management and potentially reduce the risk for clinical and subclinical metabolic disorders. To our knowledge, only Ehret *et al.* (2015) [292] predicted milk BHB based on data in the previous lactation week, using milk production traits, genomic and metabolic information. In this study, the maximum correlation coefficient (*r*) between observed and predicted milk BHB when using data in previous lactation week was 0.37, which was much lower than *r* between observed and predicted BHB using data in the same lactation week as milk BHB measurement (the highest *r* was 0.64).

Machine learning methods have been used to predict cow performance, disease, and metabolic status using large datasets [75,76]. It can be hypothesized that also hyperketonemia could be predicted with similar approaches using on-farm cow data. In fact, it is known that several on-farm cow data, like milk yield [293], milk fat and protein percent [21], and BW [20] are related to plasma BHB, or risk for clinical and subclinical ketosis. In addition, also NEI is related to plasma BHB [87], and can be expected to be a valuable variable to predict hyperketonemia, but limited available on commercial farms. The correlation among different on-farm cow variables, such as milk yield, milk fat and protein, results in intricate dependencies among explanatory variables when used together in the same model [79]. Further, it is difficult to find a specific parametric function (e.g. linear, quadratic etc.) with conventional statistic methods. In this context, analysing trough flexible algorithms in machine learning seems promising to manage large datasets with complex correlation patterns among variables [285]. We thus hypothesize that machine learning methods can predict hyperketonemia in dairy cows using on-farm cow data and NEI in early lactation. Objectives of this study were: *i*) to evaluate if hyperketonemia in dairy cows (defined as plasma BHB ≥ 1.0 mmol/L) can be predicted using on-farm cow data either in same or previous lactation week, and *ii*) to study if adding individual NEI can improve the predictive ability of the model.

7.3. Materials and Methods

The concentration of plasma BHB, on-farm cow data, and NEI of in total 424 cows originate from van Knegsel *et al.* (2007) [294], van Knegsel *et al.* (2014) [86], Chen *et al.* (2016) [88], and van Hoeij *et al.* (2017) [87]. Experimental protocols were approved by the Institutional Animal Care and Use Committee of Wageningen University. Briefly, plasma and milk samples were collected weekly. The concentration of plasma BHB was measured with kit no. RB1007 (Randox Laboratories, Ibach, Switzerland), as previously described [262]. On-farm cow data included dry period length (d), parity, BW (kg), weekly changed BW (kg/week, BW in current week minus BW in previous week), milk yield (kg/d), yield (kg/d) of milk fat, protein and lactose, the percentage of fat, protein and lactose, fat- and protein-corrected milk production (kg/d), and SCC (CHU). Net energy intake was calculated by dietary net energy concentration and feed intake of individual cows. Of all 424 individual cows, the number of cows with complete records in each lactation week are presented in Table 7.1.

Table 7.1. In lactation week 1 to 5, number of cows with complete records (on-farm cow data and individual net energy intake) in the prediction of hyperketonemia using data in same and previous lactation week [1].

Study [2]	Completed records in each lactation week				
	wk 1 to 1/2	wk 2 to 2/3	wk 3 to 3/4	wk 4 to 4/5	wk 5 to 5
van Knegsel *et al.* (2007) [294]	69/69	72/72	72/72	72/71	72
van Knegsel *et al.* (2014) [86]	58/60	58/59	91/91	91/87	91
Chen *et al.* (2016) [88]	56/59	56/56	73/73	77/76	77
van Hoeij *et al.* (2017) [126]	91/111	111/111	121/121	121/120	121
Total	274/299	297/298	357/357	361/354	361

[1] The hyperketonemia in dairy cows in lactation week 1 was not predicted by data in previous week.

[2] The experimental design, dry period length, and diet in these studies were described by van Knegsel *et al.* (2007) [294], van Knegsel *et al.* (2014) [86], Chen *et al.* (2016) [88], and van Hoeij *et al.* (2017) [126].

7.4. Results and Discussion

Preliminary analysis showed partial least square discriminant analysis (PLS-DA) outperformed random forest, artificial neural networks (ANN), and support vector machine. Therefore, PLS-DA was used to predict hyperketonemia in dairy cows in current study. Briefly, on-farm cow data and on-farm cow data combined with NEI, either in same lactation week, or in previous lactation week, were used as predictor variables (*X matrix*), whereas if dairy cows had hyperketonemia (plasma BHB $\geq$ 1.0 mmol/L [17,95]) was used as the targeted variable in classification (*Y*). Predictor variables in training and testing dataset were centered and scaled to unit variance after split in cross validation. A large proportion of the cows had a plasma BHB < 1.0 mmol/L (81.4 to 84.5% among lactation weeks), which would give too much weight to cows without hyperketonemia when training models, and would impair the accuracy when predicting hyperketonemia in dairy cows. Therefore, the data was re-sampled to obtain a reduced and balanced dataset [295]. Cows without hyperketonemia were randomly sampled to obtain a dataset where the proportion of cows without hyperketonemia was the same as the proportion of cows with hyperketonemia. After 5,000 times 5-fold cross validation, the predictive ability of models was evaluated by the accuracy (the ratio of the sum of true positive and true negative), sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). In the PLS-DA model, variable importance in projection (VIP) score was used to quantify the contribution of each variable [130]. Data pre-processing, including log transformation and scaling, model training with PLS-DA, and the determination of component number in PLS-DA were programmed in Python (version 3.6) using modules, “pandas”, “numpy”, and “sklearn”. Figures were plotted in R (version 3.3.3) with package “ggplot2”.

On average (SD), milk production was 25.5 (8.9), 33.4 (7.7), 37.2 (7.9), 39.2 (8.3), and 39.2 (8.3) kg/d for lactation week 1 to 5, respectively. Plasma BHB concentration (SD) was 0.70 (0.30), 0.83 (0.62), 0.84 (0.54), 0.84 (0.71), and 0.84 (0.72) mmol/L for lactation week 1 to 5, respectively. Incidence of hyperketonemia was 12.4%, 20.7%, 20.5%, 17.4%, and 15.8% for lactation week 1 to 5, respectively.

When applying on-farm cow data in same week to predict hyperketonemia in cows, the model predicted hyperketonemia best in week 3 (accuracy 83.8±6.6%), followed by week 5 (81.7±7.6%), week 4 (78.0±7.8%), week 1 (77.0±10.8%), and week 2

(76.0±8.6%) (Fig. 7.1-A). Through lactation week 1 to 5, sensitivity ranged from 77.3% to 83.5%, specificity from 77.2% to 86.0%, PPV from 76.6% to 84.8%, and NPV from 76.8% to 83.5% (Fig. 7.1-C).

When applying on-farm cow data in previous week to predict hyperketonemia in cows, the model predicted hyperketonemia best in week 3 (accuracy 79.0±7.9%), followed by week 5 (77.7±7.9%), week 4 (77.5±8.0%), and week 2 (72.0±8.5%) (Fig. 7.1-B). Through lactation week 1 to 4, sensitivity ranged from 72.9% to 78.7%, specificity from 73.7% to 82.7%, PPV from 73.2% to 81.3%, and NPV from 72.6% to 78.4% (Fig. 7.1-D).

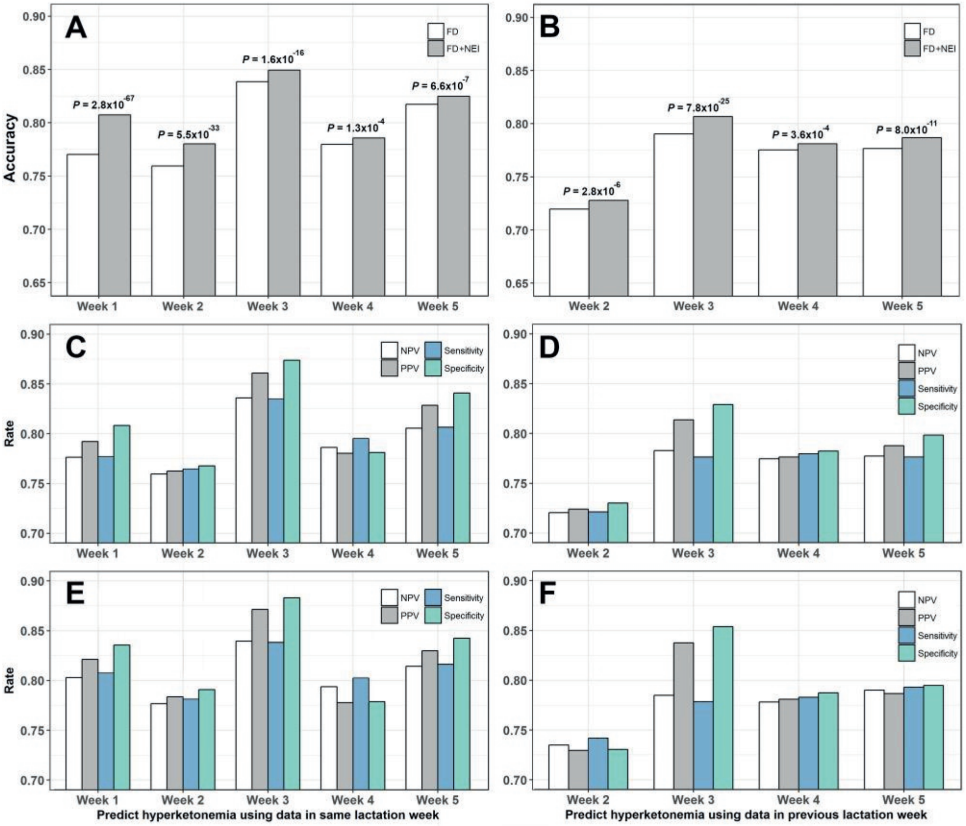


Fig. 7.1. Prediction of hyperketonemia in dairy cows using on-farm cow data (FD) and FD combined with net energy intake (NEI) by partial least squares discriminant analysis using either data in same lactation week (A: accuracy of the model per week; C (FD) and E (FD+ NEI): sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of model per week), or data in previous lactation week (B: accuracy of the model per week; D (FD) and F (FD+NEI): sensitivity, specificity, PPV, and NPV of model per week). The accuracy is defined as the ratio of the sum of true positive and true negative observations.

The ability of model to predict hyperketonemia in dairy cows differed across lactation weeks. Prediction of hyperketonemia in dairy cows in week 3 is better than the prediction in week 1, 2, 4, and 5, independent if data in same or previous lactation week were used. Dairy cows usually have a greater incidence of hyperketonemia in lactation week 2 and 3 [103,296] in early lactation. In our results, the incidence of hyperketonemia in lactation week 2 (20.5%) and 3 (20.7%) is greater than the incidence in week 1, 4, and 5 (range 12.4% to 17.4%). High incidence of higher hyperketonemia in week 3 indicates dairy cows are prone to hyperketonemia, which could be related to the maximum 83.8% accuracy to predict hyperketonemia in this week, compared with other weeks. The prediction of hyperketonemia in week 2, however, was worse than in other weeks, which could be related to the high variation of variables related to milk production at start of lactation [249]. Generally, the prediction of hyperketonemia in dairy cows was better using data in same week than data in previous week. In principle, it seems logical that on-farm cow data in the same week give a more accurate prediction because these data reflect the altered metabolic status of a cows real-time. Nevertheless, although metabolic status of cows varies highly across consecutive weeks during the first weeks of lactation [297], predictive performance based on data in the previous week is reasonable good, compared with data in same week.

When adding NEI in same week to predict hyperketonemia in cows in lactation week 1 to 5, model accuracy improved ($P < 0.05$) 3.7%, 2.1%, 1.1%, 0.6%, and 0.8%, respectively. Through lactation week 1 to 5, sensitivity ranged from 79.1% to 84.3%, specificity ranged 78.6% to 87.3%, PPV ranged 78.5% to 86.2%, and NPV ranged 78.5% to 84.2% (Fig. 7.1-E).

When adding NEI in previous week to predict hyperketonemia in cows in lactation week 2 to 5, model accuracy improved ($P < 0.05$) 0.8%, 1.6%, 0.6%, and 1.0%, respectively. Through lactation week 2 to 5, sensitivity ranged from 75.1% to 80.0%, specificity ranged 73.0% to 85.4%, PPV ranged 73.1% to 83.7%, and NPV ranged 74.1% to 79.3% (Fig. 7.1-F).

The VIP scores indicate the contribution of variables in the models to predict hyperketonemia with and without NEI (Fig. 7.2). Generally, across lactation week 1 till 5, BW change, milk fat and protein yield and milk fat and protein percentage are among the best variables predicting hyperketonemia. When NEI was added to the models to predict hyperketonemia, NEI had the highest or second highest contribution among all variables to contribute to the prediction of hyperketonemia, although the increase in model accuracy was limited. Moreover, BW [298], BW change [299] and milk fat and protein yield/percentage [300] are all related to NEI in dairy cows. This implies that part of the effect of NEI was already accounted for in the model using on-farm cow data only and explains possibly why the relative increase in accuracy when adding NEI is marginal. Further study could consider the variance of important variables not only to predict hyperketonemia, but also to predict reproductive performance [76], mastitis [85], and milk yield [81] with machine learning techniques in the context of precision dairy farming.

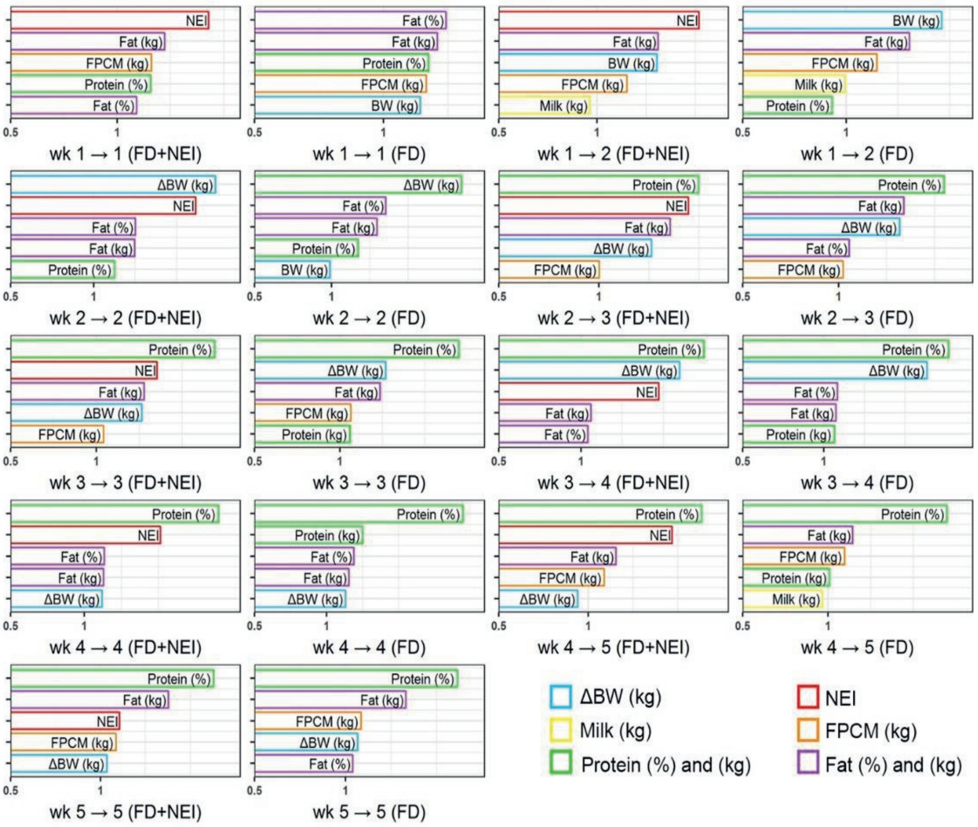


Fig. 7.2. The variable importance in projection (VIP) scores in the first principal component calculated by partial least squares discriminant analysis to predict hyperketonemia in dairy cows with on-farm cow data (FD) only and FD combined with net energy intake (NEI). The x label in each subplot formats as “wk n → m (FD or FD+NEI)”, which is presented as “Data of week n (wk n) to predict hyperketonemia in dairy cows in week m”.

Abbreviations: FPCM, fat- and protein-corrected milk production; ΔBW, changed BW; Milk, milk yield.

This study is the first report about the prediction of hyperketonemia in dairy cows using on-farm cow data and on-farm cow data combined with NEI in early lactation by PLS-DA. Besides PLS-DA, other machine learning algorithms also have been applied in studies of dairy cows to predict milk yield [81], reproductive performance [76], and metabolic status [260]. Although PLS-DA had higher accuracy than other algorithms in our preliminary analysis, it did not mean PLS-DA was always the best algorithm in practice. For more applications in precision dairy farming systems, the prediction of health and disease needs further investigation.

7.5. Conclusions

Hyperketonemia in dairy cows can be predicted by PLS-DA using on-farm cow data in both same and previous lactation week, although with some reduction in accuracy when using data in previous lactation week. Adding individual NEI improved the predictive ability of model with extra 0.6% to 3.7% accuracy. Besides NEI, changed BW, milk fat, and milk protein play important roles to predict hyperketonemia in dairy cows.

CHAPTER 8

General Discussion

8.1. Introduction

In early lactation, high-yielding dairy cows experience a negative energy balance (NEB) [86,137], which is associated with impaired health and fertility [115,301,302]. On the farm, an early and reliable screening of energy balance and metabolic status could help to identify dairy cows with an increased risk for health and fertility problems. Currently, energy balance and metabolic status can be estimated by milk characteristics, such as milk yield and milk fat [18,37,38,40], but also by the levels of ketone bodies in plasma or milk [20,22]. The current screening methods, however, are limited either by low performance to estimate energy balance when using milk characteristics, or by the invasive sampling and labour requirements in specialized laboratories when using blood plasma of dairy cows. Therefore, a non-invasive approach to estimate energy balance or metabolic status using milk samples or on-farm cow data could be interesting.

In this thesis, we used two different methods to estimate energy balance and metabolic status of dairy cows: metabolomics and machine learning techniques using on-farm cow data. Metabolomics techniques can detect small molecules present in bio-fluids [303]. The most commonly used metabolomics techniques are mass spectrometry (GC-MS for volatile molecules and LC-MS for non-volatile molecules) or proton nuclear magnetic resonance (NMR). The changes observed in the spectra measured with either NMR or MS reflect differences in amounts of metabolites present and can be used, for instance, for detection of biomarkers linked to nutrition or to disease [304]. In dairy cows, NMR and GC-MS/LC-MS studies have been applied to study differences in composition of plasma [58,92,143,305-307], milk [54,57,66,305,308,309], urine [310,311], and rumen fluid [310,312]. In our studies, using metabolomics approaches, we aimed at a better understanding of metabolic pathways related to energy balance. The outcome of these studies will make it possible to develop new biomarkers for energy balance and metabolic status as well as to develop further customized nutritional and management strategies for individual dairy cows. Machine learning, as used in this thesis, aims to develop computer algorithms that can help in the analysis of large and complex datasets [313]. In dairy cows, machine learning has been applied to estimate milk yield [81], reproductive performance [76,83], and even to identify mastitis [84,85]. Furthermore, using on-farm cow data can be expected to be important in the development of precision dairy farming applications.

The aims of this thesis were first to estimate energy balance and metabolic status of individual cows using milk metabolomics and machine learning techniques using on-farm cow data and to investigate the metabolic pathways related to energy metabolism of dairy cows in early lactation using metabolomics and proteomics techniques In this general discussion, the estimation of energy balance using plasma metabolites or milk metabolites as well the estimation of metabolic status using on-farm cow data with machine learning algorithms will be discussed. In addition, biological pathways related to energy metabolism will be discussed. Finally, conclusions and future perspectives are given.

8.2. Estimation of Energy Balance and Metabolic Status

8.2.1. Estimation of energy balance using blood and milk metabolomics data

In Chapter 2, we estimated the energy balance of 31 dairy cows in lactation week 2 and 7 using milk production traits and 52 milk metabolites obtained from (targeted) LC-MS studies. The adjusted-R^2 of the reduced models using only milk metabolites ranged from 0.65 to 0.76 (Table 2.3, Chapter 2). The adjusted-R^2 of reduced models using milk metabolites combined with milk production traits to estimate energy balance ranged from 0.83 to 0.88 (Table 2.3, Chapter 2). The combination of milk production traits with milk metabolites had a better performance to estimate energy balance than using milk production traits alone, or milk metabolites alone. In early lactation, dairy cows mobilise body reserves, such as glycogen, fat, (muscle) protein, and bone, to compensate for the nutrient deficit due to the high energy requirement for milk production [7,8], which result also in a typical increase in milk fat yield [27] and is associated with milk yield, including milk protein and lactose yield. Therefore, it can be expected that a NEB status and mobilisation of body reserves of dairy cows are reflected in both an altered metabolic status and milk yield characteristics. In our study we observe a clear relationship between milk metabolites and energy balance. Milk metabolites of cows in NEB in early lactation were attributed to cell apoptosis in the mammary gland. Concomitantly also metabolites were observed that could be attributed to cell renewal. Based on these observations we proposed that both cell apoptosis and cell renewal are paramount in dairy cows with NEB.

In Chapter 3, we reported 67 milk metabolites detected from milk samples of 87 dairy cows by the combined integration of NMR and LC-MS data. Of these 67 metabolites, 14 were positively and 25 were negatively correlated to energy balance. In line with the results of Chapter 2, where only LC-MS was used to analyse metabolic profiles of milk samples, choline, glycine, citrate, and creatinine were important metabolites related to NEB. However, of the 33 metabolites detected by NMR, 17 were observed to be modestly or highly correlated with energy balance, based on a cut-off of $r > 0.36$ or $r < -0.36$ [314]. These metabolites include acetyl-carnitine, cis-aconitate, choline, citrate, α-ketoglutarate, cytidine monophosphate (CMP), creatinine, ethanolamine, galactose-1-phosphate (Gal-1-P), N-acetyl-galactosamine (Nac-Gal), glucose-1-phosphate (Glu-1-P), glutamate, N-acetyl-glucosamine (Nac-Glu), UDP-Nac-Gal, UDP-Nac-Glu, PC and phosphocreatine.

Also NMR and LC-MS data of these 87 dairy cows can be applied to estimate the energy balance of dairy cows in week 2 of lactation (Table 8.1). Estimations are based on metabolites in milk detected from NMR (Model 1) or LC-MS (Model 2), or milk production traits (Model 3), or a combination of NMR, LC-MS, and milk production traits (Model 4, 5, 6, and 7). Model 1 estimates energy balance using 2 metabolites detected by NMR whereas model 2 estimates energy balance using 4 metabolites detected by LC-MS. Model 3 based on milk production traits has a higher estimated performance than Model 1 and Model 2, which is consistent with results presented in Chapter 2. Model 4 estimates energy balance through the combination of reliable metabolites detected from either NMR or LC-MS. Model 5 and 6 estimate energy balance with the combination of milk fat yield and milk metabolites detected either in NMR, or in LC-MS. In Model 7, milk fat yield, 3 metabolites detected by NMR and one metabolite detected by LC-MS (glycine) resulted in a model with the highest adjusted-R^2 (0.80) to estimate the energy balance of dairy cows in lactation week 2. Model 7 indicates that the combination of milk production traits and milk metabolites detected from LC-MS and NMR gives the best performance. In general, combining variables obtained from multiple sources, such as milk metabolites detected from NMR, and LC-MS, and milk production traits, improved the performance of the reduced models to estimate energy balance. Milk fat yield is important in all reduced models, and the explained variation by milk fat yield ranges from 50.9% to 84.6% in Model 3, 5, 6, and 7. In line with Chapter 2, we conclude that milk metabolites and milk production traits (in particular milk fat yield), played important roles in these models. Milk metabolites detected by NMR further improved the models as can be observed in Table 8.1.

In Chapter 4, blood plasma and milk samples of 24 dairy cows in lactation week 2 were measured using both NMR and LC-MS, and metabolic profiles of blood plasma and milk were obtained. Of 53 metabolites detected in plasma and 65 metabolites detected in milk, 24 (plasma) and 30 (milk) were correlated with energy balance of dairy cows. For example, energy balance was observed to be correlated with blood plasma metabolites BHB (r = -0.64), glycine (r = -0.81), and glucose (r = 0.80). Chapter 4 focused on the relation between metabolites present in plasma and milk, and the correlation of these metabolites with energy balance. We applied a partial least square (PLS) model to estimate energy balance of 24 dairy cows in lactation week 2 using either 53 metabolites in plasma and 65 metabolites in milk. The reason of using a PLS model, rather than using a reduced model, was to compare the performance between the blood and milk metabolic profiles for the estimation of energy balance. Based on an optimal number of PLS components, the estimation of energy balance using metabolites in plasma is 0.76 (Q^2) in a PLS model, which is higher than the estimation using metabolites detected in milk where Q^2 is 0.59 in a PLS model. These results indicate that plasma metabolites can better estimate, albeit to a limited extend, energy balance than milk metabolites.

Table 8.1. Reduced models obtained through an integrated analysis among proton nuclear magnetic resonance (NMR) spectra, liquid chromatography mass spectrometry (LC-MS) and milk production traits to estimate the energy balance (EB) of dairy cows in lactation week 2. The reduced models were selected by multivariate linear regression.

Model [1]	**No.**	**Model (Equation)**	R^2 [2]	**ad-**R^2 [3]
N [4]	**1**	EB ~ cis-Aco (87.8%) + Nac-Gal (12.2%)	0.57	0.55
M [5]	**2**	EB ~ Gly (62.5%) + carnitine (17.6%) + Met (O) (14.5%) + Arg (5.4%)	0.59	0.57
P [6]	**3**	EB = 169.1 - 514.5*fat (56.2%) + 850.0*protein (30.1%) - 19.5*milk yield (13.7%)	0.72	0.71
N+M	**4**	EB ~ Gly (41.0%) + Met (O) (20.6%) + cis-Aco (18.5%) + carnitine (15.0%) + Glu (4.8%)	0.67	0.65
N+P	**5**	EB ~ fat (84.6%) + Gal-1-P (2.3%) + Glu (5.3%) + α-KG (4.7) + creatine (3.1%)	0.79	0.77
M+P	**6**	EB ~ fat (50.9%) + Gly (22.1%) + protein (14.8%) + carnitine (7.0%) + Met (O) (5.1%)	0.78	0.76
N+M+P	**7**	EB ~ fat (69.8%) + Gly (9.6%) + Gal-1-P (9.2%) + Glu (6.8%) + UDP-Nac-Gal (4.6%)	0.81	0.80

[1] Model 1, 2, 4, 5, 6, and 7 do not show the intercept and coefficients of the variables.

[2] R^2 was obtained through 10-fold cross-validation.

[3] adjusted-R^2 (ad-R^2) considered the number of independent regressors in a model, and it was obtained through the formula, adjusted-R^2 = 1 - [(1 - R^2)(n-1)/(n-k-1)], n is the number of sample size, k is the number of independent regressors, excluding the constant.

[4] N: only milk metabolites detected by NMR are used in this model. These metabolites are marked by an asterisk.

[5] M: only milk metabolites detected by LC-MS are used in this model.

[6] P: only milk production traits are used in this model. Fat and protein represent the fat yield in milk and protein yield in milk, respectively. The unit of fat, protein, and milk yield is “kg”.

Abbreviations: cis-Aco, cis-aconitate; Arg, arginine; Gly, glycine; Gal-1-P, galactose-1-phosphate; Glu, glutamate; α-KG, α-ketoglutarate; Met (O), methionine sulfoxide; Nac-Gal, N-acetyl-galactosamine; UDP-Nac-Gal, UDP-N-acetyl-galactosamine.

Considering the relatively low differences in PLS models to estimate energy balance using metabolites observed in plasma or metabolites observed in milk we propose that milk metabolites, as milk samples are more easily obtained than blood plasma samples, could be used to estimate energy balance of dairy cows on commercial farms.

8.2.2. Generalization of the mathematical model with milk production traits

In Chapter 2, we reported the reduced model to estimate energy balance in lactation week 2 and 7 based on milk samples of 31 dairy cows (Table 2.3). In these two weeks, milk production traits, especially milk fat yield, had a high contribution to the model to estimate energy balance. The generalization of the model using only milk fat yield, however, has not been evaluated with a larger dataset than used in Chapter 2 and 3. Therefore, we applied the model using only milk fat yield to the data of 457 dairy cows in lactation week 2, 72 cows from van Knegsel *et al*. (2007) [294], 162 cows from van Knegsel *et al*. (2014) [86], 101 cows from Chen *et al*. (2016) [88], and 122 cows from van Hoeij *et al*. (2017) [126], as follows,

Model 3 (Table 2.3):

$$EB = 580.7 - 532.4 * Fat$$

Model 3 (Table 8.1):

$$EB = 169.1 - 514.5 * fat\ (56.2\%) + 850.0 * protein\ (30.1\%) - 19.5 * milk\ yield\ (13.7\%)$$

The adjusted-R^2 of Model 3 of Chapter 2 (Table 2.3) is now 0.67, which is lower than the reported performance (adjusted-R^2 = 0.78) in Chapter 2. The adjusted-R^2 of Model 3 (Table 8.1) is now 0.59, which is lower than the performance (adjusted-R^2 = 0.71) in Table 8.1). The reason could be *i*) limited sample size as only 31 cows were used in Chapter 2 or 87 cows were used in Table 8.1 leading to overfitting, *ii*) the variance among different experiments.

8.2.3. Estimation of energy balance with on-farm cow data using a stepwise regression model

In Chapter 6 and 7, variables from on-farm cow data were used to predict metabolic status lactation week 1 to 7 and hyperketonemia lactation week 1 to 5 of dairy cows. In this general discussion, we use 19 to 21 variables (depending on lactation week) from on-farm cow data to predict energy balance of dairy cows using a stepwise regression model. These variables include dry period length (d), parity, body weight (kg), body weight change by week (kg/week *i.e.* body weight in current week minus body weight in previous week), milk yield (kg/d), yield (kg/d) of milk fat, protein and lactose, the percentage of fat, protein and lactose, fat- and protein-corrected milk production (kg/d), maximum milk yield (kg/d), and minimum milk yield (kg/d). Moreover, 6 variables related to weather data were also involved in the model, including maximum ambient temperature (T, °C), minimum T (°C), mean T (°C), maximum humidity (H, %), minimum H (%), and mean H (%). These weather data were collected from www.wunderground.com, and weekly averaged data were applied in the stepwise regression model. Weather data were included because of their relation with feed intake [287,315]. The adjusted-R^2 of the reduced models with (up to) 5 variables ranges from 0.69 to 0.81 (Table 8.2).

In these models, milk production traits, such as fat- and protein-corrected milk yield (FPCM), milk fat-, protein- and lactose-related variables, were important to predict the energy balance of dairy cows. The variables important in these models in Table 8.2 were reported before in studies on the health status of dairy cows [316,317]. Besides variables related to milk production traits, such as FPCM, protein (%), and fat (%), the most important variable is body weight change (ΔBW) to estimate energy balance in lactation week 2 to 6. Body weight change was also important to estimate hyperketonemia of dairy cows in lactation week 2 to 5 (Chapter 7). The accounted variance explained by ΔBW gradually decreases from lactation week 2 (61.1%) to week 6 (7.9%), which matches with the serious body reserve mobilization in early lactation. In dairy cows, body condition score (BCS) is widely used to indicate energy balance and metabolic status [20,318], because BCS change is correlated with fat mobilization [142,319,320]. Greater prepartum fattening of cows was previously associated with a NEB status postpartum [88,321]. Body condition score, however, is not used in the current study, because BCS is not a daily obtained variable in practice. In lactation week 7, one of the weather data, minimum humidity is involved in the model, which is consistent with studies that temperature and humidity negatively affect feed intake and milk production of dairy cows [287,315,322]. Weather data could be much more important to estimate energy balance of cows in other regions, for example, India and China, compared with the Netherlands with cool summers and moderate wet winters.

Table 8.2. Reduced models to estimate the energy balance (EB) of dairy cows through lactation week (wk) 1 to 7. The reduced models were selected by stepwise regression. The maximum variable numbers in reduced model is 5.

wk [1]	Model (Equation)	R^2 [2]	ad-R^2 [3]
1	EB = 290.4 - 123.5*L_kg (53.6%) + 99.4*F_kg (38.2%) - 1.7*DPL (4.1%) + 33.4*P (2.6%) - 17.1*Lact (1.5%)	0.74	0.73
2	EB = 581.0 + 3.6*ΔBW (61.1%) - 90.1*P_kg (21.7%) + 69.1*F_kg (12.1%) - 46.9*F (4.3%) + 2.0*FPCM (0.7%)	0.81	0.81
3	EB = 395.6 - 101.5*F (37.3%) + 133.9*P (32.0%) + 2.7*ΔBW (16.3%) - 13.5*L_kg (13.1%) - 3.9*FPCM (1.4%)	0.76	0.76
4	EB = 366.7 + 156.2*P (40.7%) - 98.6*F (31.9%) + 2.6*ΔBW (16.2%) - 10.3*P_kg (8.2%) - 5.8*FPCM (3.0%)	0.78	0.78
5	EB = 622.1 - 121.0*F (45.3%) + 135.5*P (30.4%) - 13.5*MaxM (17.0%) + 1.8*ΔBW (6.7%) - 2.4*FPCM (0.6%)	0.76	0.76
6	EB = 427.6 + 150.2*P (43.4%) - 107.5*F (36.5%) - 11.4*P_kg (11.1%) + 2.1*ΔBW (7.9%) -3.5*FPCM (1.1%)	0.77	0.76
7	EB = 3.2 + 139.4*P (43.5%) - 85.8*F (36.0%) - 8.0*FPCM (10.9%) - 6.3*MaxM (6.5%) + 4.2*MinH (3.0%)	0.69	0.69
All [4]	EB = -65.6 + 43.0*wk (16.5%) - 34.7*F (2.1%) + 137.9*L (3.1%) - 18.4*FPCM (71.4%) + 4.4*ΔBW (6.9%)	0.73	0.72

[1] wk, the model with data in lactation week 1 includes 19 variables (but no ΔBW); models with data through lactation week 2 to 7 include 20 variables (ΔBW included); model including all 7 lactation weeks includes 21 variables (lactation week included).

[2] R^2 was obtained through 10-fold cross-validation.

[3] adjusted-R^2 (ad-R^2) considered the number of independent regressors in a model, and it was obtained through formula, adjusted-R^2 = $1 - [(1 - R^2)(n-1)/(n-k-1)]$, n is the number of sample size, k is the number of independent regressors, excluding the constant.

[4] model with all data through lactation week 1 to 7 includes 21 variables (lactation week included).

Abbreviations: F_kg, fat yield (kg); F, fat content (%); P_kg, protein yield (kg); P, protein content (%); L_kg, lactose yield (kg); L, lactose content (%); MaxM, maximum milk yield (kg); MinH, minimum humidity (%); FPCM, fat- and protein-corrected milk yield (kg); DPL, dry period (d); ΔBW, body weight change (kg).

8.2.4. Clustering of metabolic status of dairy cows using plasma metabolites

In Chapter 6, plasma metabolites and metabolic hormones, including free fatty acids (FFA), β-hydroxybutyrate (BHB), glucose, insulin and insulin-like growth factor 1 (IGF-1) were used to cluster dairy cows either in a good, average, or poor metabolic status. The cluster analysis can be affected by two or more metabolites which can be expected to be highly correlated (collinearity) [271,272]. In our results, the maximum correlation between two metabolites or between two metabolic hormones was 0.61, which is considered as a moderate correlation [314]. To avoid collinearity in this cluster analysis, an alternative method used is with principal components (PCs) from principal components analysis (PCA) of plasma metabolites and metabolic hormones [260,263]. The difference between the two methods is small, especially for lactation week 2, 3, and 4 (Table 8.3). Selected two PCs from the PCA could account for 65.7% to 76.1% (Fig. 6.1) of the explained variation of 5 of the plasma metabolites and metabolic hormones through lactation week 1 to 7, as discussed in Chapter 6.

Table 8.3. The result of cluster analysis with selected principal component (PC) number from principal component analysis (PCA) to 5 plasma metabolites and metabolic hormones.

Metabolic status	**Lactation week**						
	1	2	3	4	5	6	7
Cluster analysis with 5 plasma metabolites and metabolic hormones							
Good	35	80	95	129	110	118	144
Average	135	199	195	168	169	163	144
Poor	50	36	21	13	22	19	10
Cluster analysis with selected PC [1] from PCA							
Good	38	79	96	123	101	109	146
Average	124	200	192	173	177	166	146
Poor	58	36	23	14	23	25	6
Different ID of cows [2] in same clusters between two methods							
Good	5	1	1	6	9	17	4
Average	13	1	3	6	10	21	8
Poor	9	0	2	1	1	6	4

[1] The principal component number is selected byelbow rule, number is 2 through lactation week 1 to 7, as presented in *Appendix* Fig. 6.1.

[2] Different ID of cows means the number of cows with different identification between two datasets.

In Chapter 7, on-farm cow data were used to predict hyperketonemia. The threshold of plasma BHB to define hyperketonemia with data from blood plasma, as discussed in Chapter 7, was 1.0 mmol/L. The normally used thresholds to define hyperketonemia in dairy cows are for a concentration of BHB $\geq$ 1.0 mmol/L [141,323], or $\geq$ 1.2 mmol/L [56,99,105,107,289], or $\geq$ 1.4 mmol/L [96,103,104,141] in dairy cows. In our study, 58.0% to 60.9% out of 532 dairy cows had a 0-d of 30-d dry period length (DPL). When comparing this with a conventional DPL (60-d), shortening or omitting the dry period improves energy balance [137] and metabolic status [324]. With 0-d or 30-d DPL, dairy cows involved in the current study had a relatively better metabolic status, with a lower plasma concentration of BHB. The definition of 1.0 mmol/L plasma BHB resulted in a relative bigger proportion of dairy cows with hyperketonemia in the training dataset, which is 12.4%, 20.7%, 20.5%, 17.4%, and 15.8% for lactation week 1 to 5, respectively. The proportion is 6.6%, 9.1%, 10.6%, 11.4%, and 10.8%, respectively, if defining hyperketonemia as a plasma concentration of BHB ≥ 1.2 mmol/L, or 3.6%, 7.4%, 7.0%, 8.0%, and 9.1%, respectively, if defining hyperketonemia as plasma concentration of BHB ≥ 1.4 mmol/L. Therefore, we used 1.0 mmol/L plasma BHB as a threshold to define hyperketonemia of dairy cows.

8.3. Metabolic Pathways Related to Energy Balance and Metabolic Status

8.3.1. Metabolomics techniques used and metabolites detected

Nuclear magnetic resonance (NMR) is a spectroscopic method able to give a comprehensive overview of the most abundant organic compounds in a sample employing a single measurement [118]. In a biological-fluid sample, hydrogen-containing molecules will show a signal in an ^{1}H NMR spectrum, as long as they are present in concentrations above the detection limit. Alternatively, LC-MS has also been used in the detection of metabolites present in biological fluids. Metabolites can be extracted from bio-fluids and ultra-high performance liquid chromatography (UPLC) columns are used for separation of the metabolites followed by detection with a mass spectrometer. With the recent development of specific columns that can separate hydrophilic molecules, targeted approaches with Triple Quadrupole Mass Spectrometers or High Resolution Mass spectrometers are becoming into daily practice.

The NMR spectra of plasma samples were, when compared with NMR spectra of milk samples, easier to interpret. Blood, apparently, is not only a well-studied medium with regards to metabolites that can be discovered but also blood is well a well-balanced system with a specific set of abundant metabolites that can easily be assigned. In milk samples, the NMR spectrum is dominated by many resonances of lactose, which masks the signal of other metabolites, as previously described [325]. Moreover, milk is less well studied than blood samples, but a series of metabolites have been reported in previous studies using either mass spectrometry, or NMR, or both [54,125,235]. These detected metabolites in earlier studies were used for a better understanding of the metabolism of dairy cows in heat stress [57,307], ketosis [56,64], mastitis [125,326], hepatic steatosis [327], displaced abomasum [328], as well as a shortened productive lifespan [58].

8.3.2. *Body reserves mobilization and hormonal regulation*

In Chapter 4, the plasma metabolic profile of dairy cows was detected through the integration of LC-MS and NMR data. The relation of energy balance and plasma metabolites points to the process of body reserves mobilization of dairy cows in NEB [329], including muscle protein breakdown, body fat mobilization, and bone usage or breakdown.

In early lactation, dairy cows have decreased plasma glucose [142,215,216], which is consistent with our results that plasma glucose is positively correlated with energy balance (r = 0.84, Chapter 4). Glucose is endogenously produced by gluconeogenesis and glycogenolysis (the breakdown of glycogen) in the liver [221,330,331]. Propionate can contribute up to 45% in lactating cows to gluconeogenesis [332]. Propionate is produced largely in the rumen (Fig 8.1) [333,334]. In dairy cows, glucose is not only used as an energy source [217], but also used as the precursor to synthesize lactose in milk [218]. The low concentration of plasma glucose in cows in negative energy balance could indicate the high priority of the mammary gland for lactose synthesis [160,221,335].

Plasma glucose concentration is regulated by insulin, both glucose and insulin concentration are low in dairy cows in NEB [126,336]. Low plasma insulin, however, has no effect on the glucose uptake of the mammary gland, which has insulin-independent glucose transporters [220,276,337], which is beneficial for the uptake of glucose by the mammary gland during a status with low plasma insulin concentration and consequently is beneficial for milk lactose synthesis and milk production. Holstein-Friesian dairy cows have been selected through breeding programs for high milk yield resulting in a decreased plasma insulin concentration [338]. Moreover, plasma insulin plays a key role in the relationship of the growth hormone - insulin-like growth factor 1 (GH-IGF-1) axis in dairy cows in early lactation. Under physiological conditions, GH induces hepatic IGF-1 synthesis [339], and systemic IGF-1 negatively regulates GH production as a feedback [340]. In dairy cows in NEB, however, the GH-IGF-1 axis uncouples in the liver [341], which results in a reduced plasma IGF-1 concentration although plasma GH concentration is elevated [342-344]. Besides insulin, GH, and IGF-1, plasma epinephrine is also negatively correlated with energy balance (r = -0.43, Chapter 4), with higher levels of epinephrine (adrenaline) in cows with NEB.

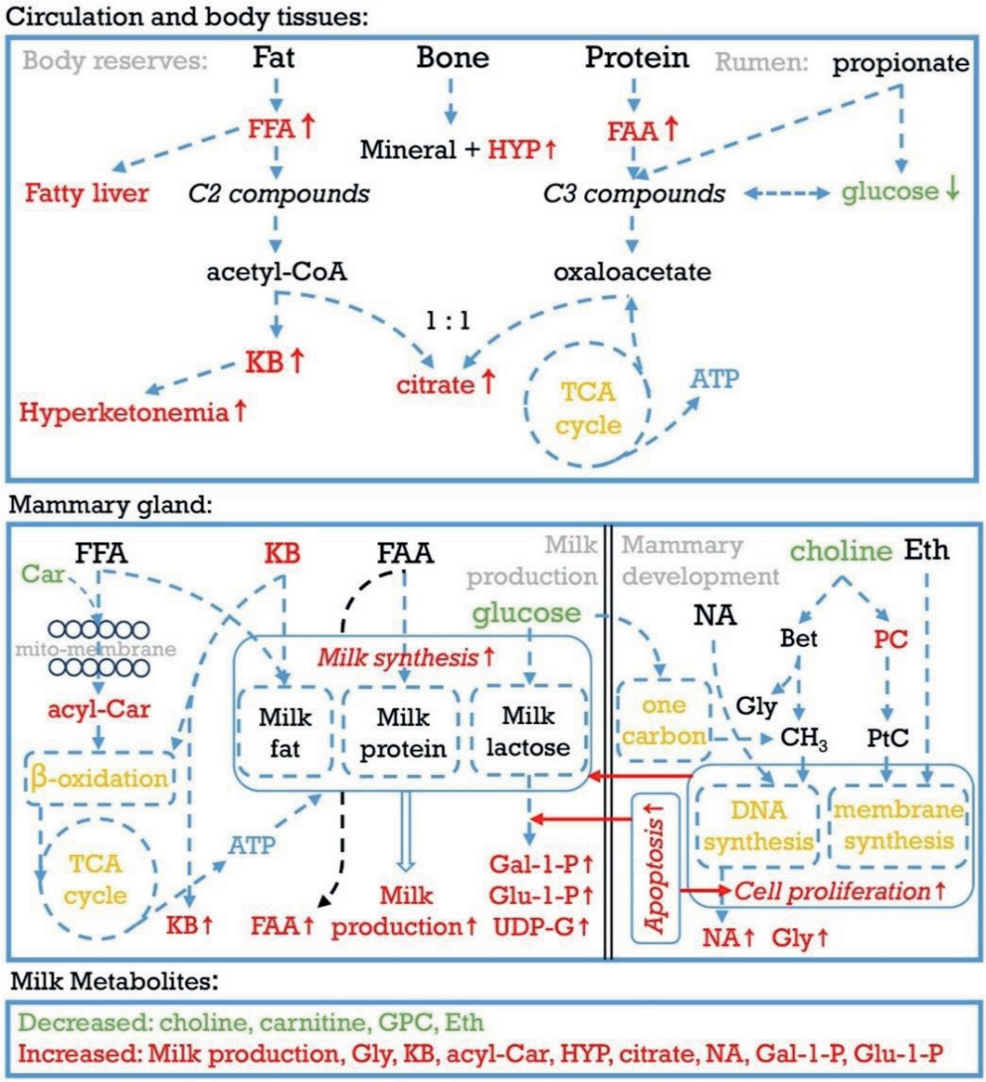


Fig. 8.1. Metabolic pathways related to energy balance of dairy cows in early lactation.

Abbreviations: Bet, betaine; Car, carnitine; Eth, ethanolamine; FAA, free amino acids; FFA, free fatty acids; Gal-1-P, galactose-1-phosphate; Glu-1-P, glucose-1-phosphate; Glu, glucose; Gly, glycine; HYP, hydroxyproline; KB, ketone bodies; NA, nucleic acids; PC, phosphocholine; PtC, phosphatidylcholine.

8.3.3. Metabolic data related to lipid metabolism

In Chapter 2, milk fat played an important role to estimate energy balance of dairy cows in lactation weeks 2 and 7. In milk, 45% to 60% of milk fat consists of medium- and short-chain fatty acids (4 to 16 carbons in length), which are mainly *de novo* synthesized in the mammary gland [345]. In the mammary gland long-chain fatty acids (LCFA) can either be secreted into milk or can be used as energy source. When fatty acids are oxidized by mitochondria for energy supply, carnitine and acyl-carnitine are important for transport of fatty acids from cytoplasm into mitochondria [346], as shown in Fig. 8.1 (See also Fig. 2.3, Chapter 2). We observed a positive correlation between energy balance and milk carnitine (Chapter 2, 3, and 4) and a negative correlation between energy balance and milk acetyl-carnitine.

Energy balance was negatively correlated with milk citrate (Chapter 2, 3, and 4). Citrate has a higher concentration in milk in early lactation, and decreases gradually as lactation progresses [347,348]. Citrate has been suggested as a marker of energy status in the dairy cow, being correlated with ketones in milk and *de novo* fatty acid synthesis [192]. Also in the current study (Chapter 4), milk citrate was positively correlated not only with acetone in milk (r = 0.45), but also with BHB in milk (r = 0.38), and milk fat yield (r = 0.74). The energy metabolism in Fig. 8.1 includes the metabolic pathways leading to the formation of the energy-carrying molecule adenosine triphosphate (ATP). Most of the ATP is created in citric acid cycle in the mitochondria, which starts from acetyl coenzyme A. Acetyl coenzyme A may be derived either from carbohydrates via glycolysis, or from fatty acids via β-oxidation, or from amino acids from protein breakdown [349].

8.3.4. Metabolic data related to lactose synthesis

Lactose is one of the main components in milk, and lactose yield of dairy cows is negatively correlated with energy balance, for example, in lactation week 2, r ranged from -0.59 to -0.77 in our studies. In the mammary gland, lactose is synthesized from two molecules of glucose in the Golgi apparatus of mammary epithelial cells (Fig. 8.2). One molecule glucose is converted into galactose that is combined with another molecule glucose to synthesize lactose [172]. The presence of Glu-1-P and Gal-1-P in milk of dairy cows in NEB as observed in our data, however, is surprising as these molecules are intermediates in the pathway for lactose synthesis. We propose that the presence of Glu-1-P and Gal-1-P in milk indicates leakage of intermediates in lactose synthesis into milk through apoptotic cells [66]. In Chapter 3 and 4, we observed that energy balance was negatively correlated with Gal-1-P (r = -0.64 and -0.57) and Glu-1-P (r = -0.51 and -0.54). This indicates that apoptosis is more severe in cows with severe NEB than in cows with a less NEB, which is in line with an earlier study who reported Gal-1-P in milk of dairy cows during NEB [66].

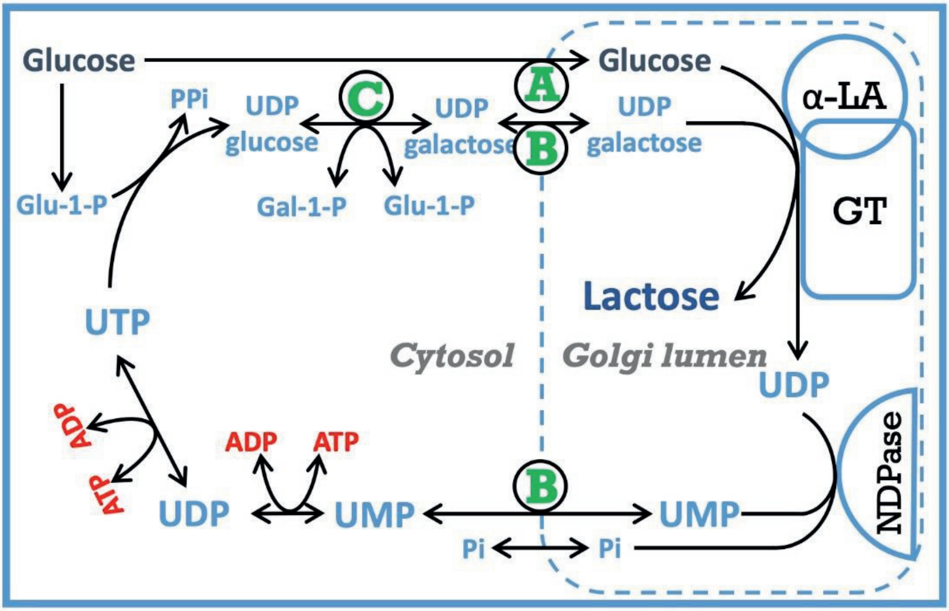


Fig. 8.2. The synthesis of milk lactose in the mammary gland of dairy cows, figure is adapted from Kuhn *et al*. (1980) [172].

Abbreviations: ADP, adenosine diphosphate; ATP, adenosine triphosphate; Glu-1-P, glucose-1-phosphate; Gal-1-P, galactose-1-phosphate; Pi, inorganic phosphate; PPi, pyrophosphate; UDP, uridine 5'-diphosphate; UMP, uridine 5'-monophosphate; UTP uridine triphosphate; GT, galactosyltransferase; α-LA, α-lactalbumin; NDPase, nucleoside diphosphatase; A, glucose transporter; B, UDP-galactose-transporter; C, galactose-1-phosphate uridylyltransferase.

8.3.5. Cell proliferation in the mammary gland

Cell apoptosis as proposed by us could occur concomitantly with cell proliferation. In a recent study it was observed that there is increase of 65% for total DNA around 10 days peripartum [180]. In Chapter 2, 3, and 4, energy balance was observed to be positively correlated with milk choline, *r* ranged from 0.63 to 0.67, and be negatively correlated with milk glycine, *r* ranged from -0.66 to -0.83. We proposed, based on these observations that choline is important as a methyl donor in the process of cell proliferation, besides its role in lipid metabolism (see Fig. 2.4, Chapter 2) [224]. Choline, folate and methionine metabolism are interrelated as all influence the production of S-adenosylmethionine, the universal donor of methyl-groups in biological reactions [147]. Besides the role of energy source and precursor of lactose (section 8.2.1 and 8.3.4), glucose could also serve as a methyl donor for one-carbon related processes in the mammary gland (Chapter 2). The role of glucose as methyl donor has been described in cancer cell proliferation [154,155]. In Chapter 2, 3, and 4, we also observed the correlation of energy balance with the intermediate of nucleic acids synthesis. For example, energy balance of dairy cows in lactation week 2 is negatively correlated with 3',5'-cyclic adenosine monophosphate (cAMP, $r = -0.43$) and cytidine monophosphate (CMP, $r = -0.58$) in Chapter 3 (Fig. 3.2). These observations support an increase in cell proliferation in cows in NEB.

A decreased glycerylphosphorylcholine/phosphocholine (GPC/PC) ratio has been observed for cancer cells and immortalized cell lines of humans [188,189,350]. In dairy cows, the decreased GPC/PC ratio was suggested as an indicator of increased risk for ketosis [64]. Here, we propose that the altered ratio of GPC/PC could also be related to the process of cell membrane synthesis during the cell proliferation in mammary gland, as shown in Fig. 3.4 (Chapter 3) [182].

8.3.6. Milk protein post-translational modifications

Post-translational modifications (PTMs), important in many biological processes, can modify the catalytic activity of eukaryote proteins [351]. In Chapter 2, 3, and 4, we observed a correlation between energy balance of dairy cows with milk choline and milk glycine. These two metabolites, glycine and choline, are both involved in the one carbon metabolism as discussed above. Choline is an important methyl donor in one carbon metabolism [147]. We therefore hypothesized that changes in choline levels could be related to changes in methylation status of proteins. The results of a proteomics study (Chapter 5) showed that specific proteins (immunoglobulin and β-lactoglobulin) have post translational methylation modifications that are also correlated with energy balance of dairy cows in week 2 of lactation. In cows in NEB these proteins were observed to have a decreased methylation as post-translational modification. In addition to changes in methylation status of some proteins, the concentration of a set of proteins (legumain, alpha-2-HS-glycoprotein, vitamin D binding protein, glycoprotein 2, folate receptor alpha, peptidyl-prolyl cis-trans isomerase B, and apolipoprotein A-IV) were correlated with energy balance. However, as our data size was limited, a more extensive study could show an even more in-depth relationship between energy balance and milk proteins modifications as well as changes in milk proteins concentration [66].

8.4. Future Studies

This thesis provides new possibilities to estimate energy balance using metabolomics data and to estimate metabolic status and energy balance based on machine learning approaches using on-farm cow data. In the context of a precision farming system, a large (and theoretically even increasing) amount of data could be collected, for example, climate conditions, body weight, milk quality variables, disease incidence, physiological data. With the increasing amount of data acquired, machine learning approaches are expected to further improve the estimation of milk quality (fat concentration, protein concentration, unsaturated versus saturated lipids), milk production, fertility, and disease incidence. In the current study, we observed that the biggest challenge using machine learning techniques is data collection. The variation obtained with different data sources is something which unfortunately often cannot be avoided and excluded. For example, the model of milk fat yield we obtained for 31 dairy cows in Chapter 2 performed worse when applied to a large population of dairy cows (section 8.2.4). When improved quality of data from different sources could be obtained machine learning could gain importance. However, the rate limiting step in this approach will be probably the local quality check of the on-farm cow data.

Machine learning approaches appear promising for applying a customized nutrition strategy to individual cows in NEB. Based on our observations, however, the main challenge in machine learning approaches will be the quality of the on-farm data. With improved standardization of on-farm data, the quality of computer-based algorithms will improve and integrated into an on-line cloud-based application for a customized nutrition strategy. Possibly also combining on-farm cow data with milk metabolomics data.

Specifically, metabolomics data can also be used to identify metabolic status of individual cows, cows at risk for metabolic disease or possibly even cows at risk for other diseases relevant for our modern high-producing dairy cows, like mastitis, or fertility-associated disorders. In the current studies, we identified a large series of important milk metabolites that are related to energy balance of dairy cows in early lactation, *i.e.* choline, citrate, cytidine monophosphate, glucose-1-phosphate, galactose-1-phosphate, cis-aconitate, fucose, N-acetyl-galactosamine (Nac-Gal), N-acetyl-glucosamine (Nac-Glu), UDP-Nac-Gal, UDP-Nac-Glu, glycine, and carnitine. In the future, these metabolites could be detected using lab-on-a chip systems for integrated bioanalyses. The estimation of energy balance and metabolic status, and in the future maybe also other physiological statuses, based on milk metabolites could customise management and nutrition strategies for individual cows in NEB. Based on nutritional perspectives, for example, extra supply of glucose or glucogenic diets to individual cows could help for cows in NEB [294], and extra supply of rumen-protected choline to individual cows could help to overcome the limitations of the one carbon metabolism at low choline levels [352,353].

8.5. Conclusions

We applied two different methods to estimate energy balance of dairy cows, a machine learning approach using on-farm cow data or metabolomics techniques with a reduced model approach. The results of the machine learning approach appeared promising using limited datasets but when extending the dataset size the performance dropped, which indicate that either data collection, data or variable selection, or the models themselves should be further optimized. Part of this optimization could be realized by an integration of machine learning techniques with milk metabolomics data. Metabolomics techniques can be more easily standardized than machine learning techniques but the challenge with metabolomics approaches is that specialized instrumentation and highly skilled technical people are needed for data acquisition. With developments of lab on chip devices to determine the amount of specific metabolites (biomarkers) in milk samples less emphasis on specialized instrumentation or skilled technical personal is necessary. Based on our studies as described in this thesis several important biomarkers could be used for detection of energy status on a cow in early lactation: choline, citrate, galactose-1-phosphate, glucose-1-phosphate, glycine, phosphocholine, UDP-N-acetyl-galactosamine, and UDP-N-acetyl-glucosamine. When the amounts of these metabolites or a combination of these metabolites will be detected using lab on a chip system, the farmer can decide to supplement specific compounds (glucose, rumen protected choline) to the cow in order to alleviate the problems of an animal in NEB.

Reference list

1 CRV. **2015**. Annual report 2015. (CRV).

2 USDA-NASS. (2015).

3 **Olori, V. E. *et al.*** (1999). Fit of standard models of the lactation curve to weekly records of milk production of cows in a single herd. *Livest. Prod. Sci.* 58, 55-63.

4 **Stanton, T. *et al.*** (1992). Estimating milk, fat, and protein lactation curves with a test day model. *J. Dairy Sci.* 75, 1691-1700.

5 **Garnsworthy, P. *et al.*** (1982). The effect of body condition of dairy cows at calving on their food intake and performance when given complete diets. *Anim. Sci.* 35, 113-119.

6 **Garnsworthy, P. C. *et al.*** (1987). The influence of body condition at calving and dietary protein supply on voluntary food intake and performance in dairy cows. *Anim. Sci.* 44, 347-353.

7 **Grummer, R. R.** (1995). Impact of changes in organic nutrient metabolism on feeding the transition dairy cow. *J. Anim. Sci.* 73, 2820-2833.

8 **Collard, B. *et al.*** (2000). Relationships between energy balance and health traits of dairy cattle in early lactation. *J. Dairy Sci.* 83, 2683-2690.

9 **Grummer, R. R.** (1993). Etiology of lipid-related metabolic disorders in periparturient dairy cows. *J. Dairy Sci.* 76, 3882-3896.

10 **Mostafavi, M. *et al.*** (2015). Evaluation of fructosamine as a new biomarker for diagnosis of hepatic lipidosis in dairy cows. *Anim. Prod. Sci.* 55, 1005-1010.

11 **Herdt, T. H.** (2000). Ruminant Adaptation to Negative Energy Balance: Influences on the Etiology of Ketosis and Fatty Liver. *Vet. Clin. North Am. Food Anim. Pract.* 16, 215-230.

12 **McCarthy, S. D. *et al.*** (2010). Negative energy balance and hepatic gene expression patterns in high-yielding dairy cows during the early postpartum period: a global approach. *Physiol. Genomics* 42, 188-199.

13 **Wathes, D. *et al.*** (2007). Influence of negative energy balance on cyclicity and fertility in the high producing dairy cow. *Theriogenology* 68, S232-S241.

14 **Patton, J. *et al.*** (2007). Relationships among milk production, energy balance, plasma analytes, and reproduction in Holstein-Friesian cows. *J. Dairy Sci.* 90, 649-658.

15 **Mayasari, N. *et al.*** (2016). The effects of dry period length and dietary energy source on natural antibody titers and mammary health in dairy cows. *Prev. Vet. Med.* 127, 1-9.

16 **van Knegsel, A. *et al.*** (2007). Natural antibodies related to energy balance in early lactation dairy cows. *J. Dairy Sci.* 90, 5490-5498.

17 **Seifi, H. A. *et al.*** (2011). Metabolic predictors of post-partum disease and culling risk in dairy cattle. *Vet. J.* 188, 216-220.

18 **Friggens, N. *et al.*** (2007). On the use of milk composition measures to predict the energy balance of dairy cows. *J. Dairy Sci.* 90, 5453-5467.

19 **Mäntysaari, P. *et al.*** (2010). Predicting early lactation energy balance in primiparous Red Dairy Cattle using milk and body traits. *Acta Agric. Scand. A Anim. Sci.* 60, 79-87.

20 **Bernabucci, U. *et al.*** (2005). Influence of body condition score on relationships between metabolic status and oxidative stress in periparturient dairy cows. *J. Dairy Sci.* 88, 2017-2026.

21 **Duffield, T. F. *et al.*** (1997). Use of test day milk fat and milk protein to detect subclinical ketosis in dairy cattle in Ontario. *Can. Vet. J.* 38, 713.

22 **Enjalbert, F. *et al.*** (2001). Ketone bodies in milk and blood of dairy cows: Relationship between concentrations and utilization for detection of subclinical ketosis. *J. Dairy Sci.* 84, 583-589.

23 **Løvendahl, P. *et al.*** (2010). Limits to prediction of energy balance from milk composition measures at individual cow level. *J. Dairy Sci.* 93, 1998-2006.

24 **Andersson, L.** (1984). Concentrations of blood and milk ketone bodies, blood isopropanol and plasma glucose in dairy cows in relation to the degree of hyperketonaemia and clinical signs. *Zentralblatt für Veterinärmedizin Reihe A* 31, 683-693.

25 **Stewart, M. *et al.*** (2007). Non-invasive measurement of stress in dairy cows using infrared thermography. *Physiol. Behav.* 92, 520-525.

26 **Alam, M. *et al.*** (1986). Effect of various veterinary procedures on plasma concentrations of cortisol, luteinising hormone and prostaglandin F2 alpha metabolite in the cow. *Vet. Rec.* 118, 7-10.

27 **de Vries, M. J. *et al.*** (2000). Energy balance of dairy cattle in relation to milk production variables and fertility. *J. Dairy Sci.* 83, 62-69.

28 **Bell, A. W.** (1995). Regulation of organic nutrient metabolism during transition from late pregnancy to early lactation. *J. Anim. Sci.* 73, 2804-2819.

29 **Tamminga, S. *et al.*** (1997). Changes in composition and energy content of liveweight loss in dairy cows with time after parturition. *Livest. Prod. Sci.* 52, 31-38.

30 **Komaragiri, M. V. *et al.*** (1997). Factors affecting body tissue mobilization in early lactation dairy cows. 1. Effect of dietary protein on mobilization of body fat and Protein1. *J. Dairy Sci.* 80, 929-937.

31 **Hoffman, K. *et al.*** (2000). Prescribed grazing and feeding management of lactating dairy cows. *New York State Grazing Lands Conservation Initiative in Cooperation with the USDA-NRCS, Syracuse, New York*.

32 **Schröder, U. J. *et al.*** (2006). Invited Review: Methods to Determine Body Fat Reserves in the Dairy Cow with Special Regard to Ultrasonographic Measurement of Backfat Thickness. *J. Dairy Sci.* 89, 1-14.

33 **van Es, A.** (1975). Feed evaluation for dairy cows. *Livest. Prod. Sci.* 2, 95-107.

34 Friggens, N. *et al.* in *Boologisk Selskabs Seminar, Dansk Boologisk Selskab.* 56-61.

35 **Heuer, C. *et al.*** (2001). Prediction of energy balance in high yielding dairy cows with test-day information. *J. Dairy Sci.* 84, 471-481.

36 **McParland, S. *et al.*** (2011). The use of mid-infrared spectrometry to predict body energy status of Holstein cows1. *J. Dairy Sci.* 94, 3651-3661.

37 **Grieve, D. *et al.*** (1986). Relationship between milk composition and some nutritional parameters in early lactation. *Livest. Prod. Sci.* 14, 239-254.

38 **Heuer, C. *et al.*** (2000). Prediction of energy balance in a high yielding dairy herd in early lactation: model development and precision. *Livest. Prod. Sci.* 65, 91-105.

39 **Reist, M. *et al.*** (2002). Estimation of energy balance at the individual and herd level using blood and milk traits in high-yielding dairy cows. *J. Dairy Sci.* 85, 3314-3327.

40 **Alphonsus, C. *et al.*** (2015). Prediction of Energy Balance of Friesian x Bunaji dairy Cows using Multivariate Milk Composition Models. *Global J. Anim. Sci. Res.* 3, 628-635.

41 **Stoop, W. *et al.*** (2009). Effect of lactation stage and energy status on milk fat composition of Holstein-Friesian cows. *J. Dairy Sci.* 92, 1469-1478.

42 **Gross, J. *et al.*** (2011). Milk fatty acid profile related to energy balance in dairy cows. *J. Dairy Res.* 78, 479-488.

43 **Ducháček, J. *et al.*** (2013). Changes in milk fatty acid composition in relation to indicators of energy balance in Holstein cows. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis* 60, 29-38.

44 **Gustafsson, A. *et al.*** (1996). Milk acetone concentration as an indicator of hyperketonaemia in dairy cows: the critical value revised. *Anim. Sci.* 63, 183-188.

45 **Wathes, D. *et al.*** (2007). Multiple correlation analyses of metabolic and endocrine profiles with fertility in primiparous and multiparous cows. *J. Dairy Sci.* 90, 1310-1325.

46 **Oetzel, G. R.** (2004). Monitoring and testing dairy herds for metabolic disease. *Veterinary Clinics: Food Animal Practice* 20, 651-674.

47 **Schulz, K. *et al.*** (2014). Effects of prepartal body condition score and peripartal energy supply of dairy cows on postpartal lipolysis, energy balance and ketogenesis: an animal model to investigate subclinical ketosis. *J. Dairy Res.* 81, 257-266.

48 **Heuer, C. *et al.*** (1999). Postpartum body condition score and results from the first test day milk as predictors of disease, fertility, yield, and culling in commercial dairy herds. *J. Dairy Sci.* 82, 295-304.

49 **Simensen, E. *et al.*** (1990). Ketosis treatment and milk yield in dairy cows related to milk acetoacetate levels. *Acta Vet. Scand.* 31, 433-440.

50 **González, L. *et al.*** (2008). Changes in feeding behavior as possible indicators for the automatic monitoring of health disorders in dairy cows. *J. Dairy Sci.* 91, 1017-1028.

51 **Hof, G. *et al.*** (1997). Milk urea nitrogen as a tool to monitor the protein nutrition of dairy cows. *J. Dairy Sci.* 80, 3333-3340.

52 **Roseler, D. *et al.*** (1993). Dietary protein degradability effects on plasma and milk urea nitrogen and milk nonprotein nitrogen in Holstein cows. *J. Dairy Sci.* 76, 525-534.

53 **Schepers, A. *et al.*** (1998). Evaluation of the utilization of dietary nitrogen by dairy cows based on urea concentration in milk. *J. Dairy Sci.* 81, 579-584.

54 **Klein, M. S. *et al.*** (2010). Nuclear magnetic resonance and mass spectrometry-based milk metabolomics in dairy cows during early and late lactation. *J. Dairy Sci.* 93, 1539-1550.

55 **Zhang, H. *et al.*** (2013). Plasma metabolomic profiling of dairy cows affected with ketosis using gas chromatography/mass spectrometry. *BMC Vet. Res.* 9, 1.

56 **Sun, L. *et al.*** (2014). 1H-Nuclear magnetic resonance-based plasma metabolic profiling of dairy cows with clinical and subclinical ketosis. *J. Dairy Sci.* 97, 1552-1562.

57 **Tian, H. *et al.*** (2016). Integrated Metabolomics Study of the Milk of Heat-stressed Lactating Dairy Cows. *Sci, Rep.* 6, 24208.

58 **Huber, K. *et al.*** (2016). Metabotypes with properly functioning mitochondria and anti-inflammation predict extended productive life span in dairy cows. *Sci. Rep.* 6, 24642.

59 **Xu, C. *et al.*** (2016). 1h-nuclear magnetic resonance-based plasma metabolic profiling of dairy cows with fatty liver. *Asian-Austral. J. Anim. Sci.* 29, 219.

60 **Butler, S. T. *et al.*** (2003). Insulin restores GH responsiveness during lactation-induced negative energy balance in dairy cattle: effects on expression of IGF-I and GH receptor 1A. *J. Endocrinol.* 176, 205-217.

61 **Ryan, D. *et al.*** (2006). Metabolomics: the greatest omics of them all? *Anal. Chim.* 78, 7954-7958.

62 **Wang, X. *et al.*** (2011). Pattern recognition approaches and computational systems tools for ultra performance liquid chromatography–mass spectrometry-based comprehensive metabolomic profiling and pathways analysis of biological data sets. *Anal. Chem.* 84, 428-439.

63 **Xi, X. *et al.*** (2017). Ultra-performance liquid chromatography-quadrupole-time of flight mass spectrometry MSE-based untargeted milk metabolomics in dairy cows with subclinical or clinical mastitis. *J. Dairy Sci.* 100, 4884-4896.

64 **Klein, M. S. *et al.*** (2012). NMR metabolomic analysis of dairy cows reveals milk glycerophosphocholine to phosphocholine ratio as prognostic biomarker for risk of ketosis. *J. Proteome Res.* 11, 1373-1381.

65 **Garnsworthy, P. C. *et al.*** (2006). Variation of Milk Citrate with Stage of Lactation and De Novo Fatty Acid Synthesis in Dairy Cows. *J. Dairy Sci.* 89, 1604-1612.

66 **Lu, J. *et al.*** (2013). Changes in milk proteome and metabolome associated with dry period length, energy balance, and lactation stage in postparturient dairy cows. *J. Proteome Res.* 12, 3288-3296.

67 **O'Donnell, R. *et al.*** (2004). Milk proteomics. *Int. Dairy J.* 14, 1013-1023.

68 Fox, P. F. *et al.* Dairy chemistry and biochemistry. (Springer, 1998).

69 **Reinhardt, T. A. *et al.*** (2006). Bovine milk fat globule membrane proteome. *J. Dairy Res.* 73, 406-416.

70 **Eigel, W. *et al.*** (1984). Nomenclature of proteins of cow's milk: Fifth revision1. *J. Dairy Sci.* 67, 1599-1631.

71 **Lu, J. *et al.*** (2014). Identification of lipid synthesis and secretion proteins in bovine milk. *J. Dairy Res.* 81, 65-72.

72 **Johnston, D. *et al.*** (2018). Differences in the bovine milk whey proteome between early pregnancy and the estrous cycle. *Theriogenology* 114, 301-307.

73 **Hogarth, C. J. *et al.*** (2004). Differential protein composition of bovine whey: a comparison of whey from healthy animals and from those with clinical mastitis. *Proteomics* 4, 2094-2100.

74 **Boehmer, J. *et al.*** (2010). Proteomic analysis of the temporal expression of bovine milk proteins during coliform mastitis and label-free relative quantification. *J. Dairy Sci.* 93, 593-603.

75 **Shahinfar, S. *et al.*** (2012). Prediction of breeding values for dairy cattle using artificial neural networks and neuro-fuzzy systems. *Comput. Math. Method. M.* 2012.

76 **Hempstalk, K. *et al.*** (2015). Machine learning algorithms for the prediction of conception success to a given insemination in lactating dairy cows. *J. Dairy Sci.* 98, 5262-5273.

77 **Bauman, D. E. *et al.*** (2003). Nutritional regulation of milk fat synthesis. *Annu. Rev. Nutr.* 23, 203-227.

78 **German, J. B. *et al.*** (1997). Milk fat composition: Targets for alteration of function and nutrition.

79 **Wilmink, J.** (1987). Adjustment of test-day milk, fat and protein yield for age, season and stage of lactation. *Livest. Prod. Sci.* 16, 335-348.

80 **Dormann, C. F. *et al.*** (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36, 27-46.

81 **Gianola, D. *et al.*** (2011). Predicting complex quantitative traits with Bayesian neural networks: a case study with Jersey cows and wheat. *BMC Genet.* 12, 87.

82 **Borchers, M. *et al.*** (2017). Machine-learning-based calving prediction from activity, lying, and ruminating behaviors in dairy cattle. *J. Dairy Sci.* 100, 5664-5674.

83 **Shahinfar, S. *et al.*** (2014). Prediction of insemination outcomes in Holstein dairy cattle using alternative machine learning algorithms. *J. Dairy Sci.* 97, 731-742.

84 **Cavero, D. *et al.*** (2008). Mastitis detection in dairy cows by application of neural networks. *Livest. Sci.* 114, 280-286.

85 **Sun, Z. *et al.*** (2010). Detection of mastitis and its stage of progression by automatic milking systems using artificial neural networks. *J. Dairy Res.* 77, 168-175.

86 **van Knegsel, A. *et al.*** (2014). Effect of dry period length and dietary energy source on energy balance, milk yield, and milk composition of dairy cows. *J. Dairy Sci.* 97, 1499-1512.

87 **van Hoeij, R. J. *et al.*** (2017). Consequences of dietary energy source and energy level on energy balance, lactogenic hormones, and lactation curve characteristics of cows after a short or omitted dry period. *J. Dairy Sci.* 100, 8544-8564.

88 **Chen, J. *et al.*** (2016). Effects of dry period length and dietary energy source on milk yield, energy balance, and metabolic status of dairy cows over 2 consecutive years: Effects in the second year. *J. Dairy Sci.* 99, 4826-4838.

89 **Acorda, J. *et al.*** (1995). Comparative evaluation of fatty infiltration of the liver in dairy cattle by using blood and serum analysis, ultrasonography, and digital analysis. *Vet. Quart.* 17, 12-14.

90 **Starke, A. *et al.*** (2010). Noninvasive detection of hepatic lipidosis in dairy cows with calibrated ultrasonographic image analysis. *J. Dairy Sci.* 93, 2952-2965.

91 **Mostafavi, M. *et al.*** (2013). Optimal thresholds of metabolic indicators of hepatic lipidosis in dairy cows. *Rev. Med. Vet.* 164.

92 **Imhasly, S. *et al.*** (2014). Metabolomic biomarkers correlating with hepatic lipidosis in dairy cows. *BMC Vet. Res.* 10, 1.

93 **Filar, J.** (1979). Über den Gehalt an β-Hydroxybutyrat, Azetazetat und Azeton im Blut von gesunden und an Ketose erkrankten Kühen. *Wien. Tieraerztl. Monatsschr.* 66, 377-380.

94 **Erhardt, G. *et al.*** (1982). Changes in concentration of citrate in milk of cows investigated peripartally, during lactation, and after experimental infections of the

mammary gland, and relations to milk constituents. *Milchwissenschaft (Germany, FR)*.

95 **Whitaker, D. *et al.*** (1983). Subclinical ketosis and serum beta-hydroxybutyrate levels in dairy cattle. *Br. Vet. J.* 139, 462-463.

96 **Nielen, M. *et al.*** (1994). Evaluation of two cowside tests for the detection of subclinical ketosis in dairy cows. *Can. Vet. J.* 35, 229.

97 **Dirksen, G. *et al.*** (1995). Ketosediagnostik: Semiquantitative Bestimmung von Beta-Hydroxybuttersäure in Kuhmilch mit einem neuen Trockentest. *Tierärztliche Umschau* 50, 239-244.

98 **Steen, A. *et al.*** (1996). Evaluation of additional acetone and urea analyses, and of the fat-lactose-quotient in cow milk samples in the herd recording system in Norway. *J. Vet. Med. A* 43, 181-191.

99 **Geishauser, T. *et al.*** (1998). Evaluation of five cowside tests for use with milk to detect subclinical ketosis in dairy cows. *J. Dairy Sci.* 81, 438-443.

100 Hünniger, F. **1998**. Einsatz der Milchacetonbestimmung bei der Betreuung von Milchviehherden, uitgever niet vastgesteld.

101 **Hansen, P. W.** (1999). Screening of dairy cows for ketosis by use of infrared spectroscopy and multivariate calibration. *J. Dairy Sci.* 82, 2005-2010.

102 **Dobbelaar, P. *et al.*** (1998). Detectie van hyperketonemie met behulp van beoordeling van de melkproductie en twee melktesten. *Tijdschrift voor diergeneeskunde* 123, 145-150.

103 **Geishauser, T. *et al.*** (2000). Evaluation of eight cow-side ketone tests in milk for detection of subclinical ketosis in dairy cows. *J. Dairy Sci.* 83, 296-299.

104 **Carrier, J. *et al.*** (2004). Evaluation and use of three cowside tests for detection of subclinical ketosis in early postpartum cows. *J. Dairy Sci.* 87, 3725-3735.

105 **Clark, C. E. F. *et al.*** (2005). The use of indicators to assess the degree of mobilisation of body reserves in dairy cows in early lactation on a pasture-based diet. *Livest. Prod. Sci.* 94, 199-211.

106 **de Roos, A. P. W. *et al.*** (2007). Screening for Subclinical Ketosis in Dairy Cattle by Fourier Transform Infrared Spectrometry. *J. Dairy Sci.* 90, 1761-1766.

107 **Van Haelst, Y. *et al.*** (2008). Elevated concentrations of oleic acid and long-chain fatty acids in milk fat of multiparous subclinical ketotic cows. *J. Dairy Sci.* 91, 4683-4686.

108 **Goldhawk, C. *et al.*** (2009). Prepartum feeding behavior is an early indicator of subclinical ketosis. *J. Dairy Sci.* 92, 4971-4977.

109 **Asl, A. N. *et al.*** (2011). Prevalence of subclinical ketosis in dairy cattle in the Southwestern Iran and detection of cutoff point for NEFA and glucose concentrations for diagnosis of subclinical ketosis. *Prev. Vet. Med.* 100, 38-43.

110 **Vanholder, T. *et al.*** (2015). Risk factors for subclinical and clinical ketosis and association with production parameters in dairy cows in the Netherlands. *J. Dairy Sci.* 98, 880-888.

111 **Xu, C. *et al.*** (2015). Mass spectral analysis of urine proteomic profiles of dairy cows suffering from clinical ketosis. *Vet. Quart.* 35, 133-141.

112 **Mann, S. *et al.*** (2016). Short communication: Association of milk fatty acids with early lactation hyperketonemia and elevated concentration of nonesterified fatty acids. *J. Dairy Sci.* 99, 5851-5857.

113 **Zhang, G. *et al.*** (2017). Metabotyping reveals distinct metabolic alterations in ketotic cows and identifies early predictive serum biomarkers for the risk of disease. *Metabolomics* 13, 43.

114 **Kokkonen, T. *et al.*** (2005). Effect of body fatness and glucogenic supplement on lipid and protein mobilization and plasma leptin in dairy cows. *J. Dairy Sci.* 88, 1127-1141.

115 **Grummer, R. R.** (1993). Etiology of Lipid-Related Metabolic Disorders in Periparturient Dairy Cows. *J. Dairy Sci.* 76, 3882-3896.

116 **Butler, W. R. *et al.*** (1989). Interrelationships between Energy-Balance and Postpartum Reproductive Function in Dairy-Cattle. *J. Dairy Sci.* 72, 767-783.

117 NRC. **2001**. Nutrient Requirements of Dairy Cattle. (National Academies Press).

118 **Nicholson, J. K. *et al.*** (2008). Systems biology: metabonomics. *Nature* 455, 1054-1056.

119 **Fernie, A. R. *et al.*** (2004). Metabolite profiling: from diagnostics to systems biology. *Nat. Rev. Mol. Cell Biol.* 5, 763-769.

120 **Saccenti, E. *et al.*** (2014). Of monkeys and men: A metabolomic analysis of static and dynamic urinary metabolic phenotypes in two species. *PloS one* 9, e106077.

121 **Silanikove, N. *et al.*** (2014). Milk metabolites as indicators of mammary gland functions and milk quality. *J. Dairy Res.* 81, 358-363.

122 **Pyörälä, S.** (2003). Indicators of inflammation in the diagnosis of mastitis. *Vet. Res.* 34, 565-578.

123 **Chagunda, M. G. *et al.*** (2006). L-lactate dehydrogenase and N-acetyl-β-D-glucosaminidase activities in bovine milk as indicators of non-specific mastitis. *J. Dairy Res.* 73, 431-440.

124 **Hettinga, K. *et al.*** (2009). The origin of the volatile metabolites found in mastitis milk. *Vet. Microbiol.* 137, 384-387.

125 **Sundekilde, U. K. *et al.*** (2013). Nuclear magnetic resonance metabonomics reveals strong association between milk metabolites and somatic cell count in bovine milk. *J. Dairy Sci.* 96, 290-299.

126 **van Hoeij, R. J. *et al.*** (2017). The effect of dry period length and postpartum level of concentrate on milk production, energy balance, and plasma metabolites of dairy cows across the dry period and in early lactation. *J. Dairy Sci.* 100, 5863-5879.

127 **Suzuki, M. *et al.*** (2016). LC–MS/MS-based metabolome analysis detected changes in the metabolic profiles of small and large intestinal adenomatous polyps in Apc Min/+ mice. *Metabolomics* 12, 68.

128 **Matsubara, A. *et al.*** (2014). Supercritical fluid extraction as a preparation method for mass spectrometry of dried blood spots. *J. Chromatogr. B* 969, 199-204.

129 **Szymanska, E. *et al.*** (2012). Double-check: validation of diagnostic statistics for PLS-DA models in metabolomics studies. *Metabolomics* 8, S3-S16.

130 **Wold, S. *et al.*** (2001). PLS-regression: a basic tool of chemometrics. *Chemometrics Intellig. Lab. Syst.* 58, 109-130.

131 **Dejean, S. *et al.*** (2013). MixOmics: Omics data integration project. *R Package*.

132 Varmuza, K. *et al.* **2016**. Introduction to multivariate statistical analysis in chemometrics. (CRC press).

133 Maindonald, J. *et al.* **2006**. Data analysis and graphics using R: an example-based approach. (Cambridge University Press).

134 **Linzell, J. *et al.*** (1971). Mechanism of milk secretion. *Physiol. Rev.* 51, 564-597.

135 **Sundekilde, U. K. *et al.*** (2011). Relationship between the metabolite profile and technological properties of bovine milk from two dairy breeds elucidated by NMR-based metabolomics. *J. Agric. Food Chem.* 59, 7360-7367.

136 **Andersen, J. B. *et al.*** (2005). The effects of dry period versus continuous lactation on metabolic status and performance in periparturient cows. *J. Dairy Sci.* 88, 3530-3541.

137 **Rastani, R. *et al.*** (2005). Reducing dry period length to simplify feeding transition cows: Milk production, energy balance, and metabolic profiles. *J. Dairy Sci.* 88, 1004-1014.

138 **Coulon, J. *et al.*** (1991). Variations in milk output and milk protein content in response to the level of energy supply to the dairy cow: a review. *Livest. Prod. Sci.* 29, 31-47.

139 **Nocek, J. *et al.*** (1988). Protein and energy as an integrated system. Relationship of ruminal protein and carbohydrate availability to microbial synthesis and milk production. *J. Dairy Sci.* 71, 2070-2107.

140 **Buttchereit, N. *et al.*** (2010). Evaluation of five lactation curve models fitted for fat: protein ratio of milk and daily energy balance. *J. Dairy Sci.* 93, 1702-1712.

141 **Walsh, R. B. *et al.*** (2007). The Effect of Subclinical Ketosis in Early Lactation on Reproductive Performance of Postpartum Dairy Cows. *J. Dairy Sci.* 90, 2788-2796.

142 **Weber, C. *et al.*** (2013). Variation in fat mobilization during early lactation differently affects feed intake, body condition, and lipid and glucose metabolism in high-yielding dairy cows. *J. Dairy Sci.* 96, 165-180.

143 **Klein, M. S. *et al.*** (2013). Correlations between milk and plasma levels of amino and carboxylic acids in dairy cows. *J. Proteome Res.* 12, 5223-5232.

144 **Meijer, G. *et al.*** (1995). Free amino acids in plasma and muscle of high yielding dairy cows in early lactation. *J. Dairy Sci.* 78, 1131-1141.

145 **Friesen, R. W. *et al.*** (2007). Relationship of dimethylglycine, choline, and betaine with oxoproline in plasma of pregnant women and their newborn infants. *J. Nutr.* 137, 2641-2646.

146 Institute of Medicine (US) Standing Committee on the Scientific Evaluation of Dietary Reference Intakes and its Panel on Folate, O. B. V., and Choline. **1998**. Dietary reference intakes for thiamin, riboflavin, niacin, vitamin B6, folate, vitamin B12, pantothenic acid, biotin, and choline. (National Academies Press).

147 Grummer, R. in *Proc. Cornell Nutr. Conf. Cornell University, Syracuse, NY.* 22-27.

148 **Pinotti, L. *et al.*** (2005). Choline: Is there a need in the lactating dairy cow? *Livest. Prod. Sci.* 98, 149-152.

149 **Ardalan, M. *et al.*** (2010). Effect of rumen-protected choline and methionine on physiological and metabolic disorders and reproductive indices of dairy cows. *J. Anim. Physiol. Anim. Nutr.* 94.

150 **Cooke, R. *et al.*** (2007). Supplemental choline for prevention and alleviation of fatty liver in dairy cattle. *J. Dairy Sci.* 90, 2413-2418.

151 **Lima, F. S. *et al.*** (2012). Effects of feeding rumen-protected choline on incidence of diseases and reproduction of dairy cows. *Vet. J.* 193, 140-145.

152 **Servillo, L. *et al.*** (2014). Where Does N$^{\varepsilon}$-Trimethyllysine for the Carnitine Biosynthesis in Mammals Come from? *PloS one* 9, e84589.

153 **Giesbertz, P. *et al.*** (2015). An LC-MS/MS method to quantify acylcarnitine species including isomeric and odd-numbered forms in plasma and tissues. *J. Lipid Res.*, jlr. D061721.

154 **Labuschagne, C. F. *et al.*** (2014). Serine, but not glycine, supports one-carbon metabolism and proliferation of cancer cells. *Cell Rep.* 7, 1248-1258.

155 **Amelio, I. *et al.*** (2014). Serine and glycine metabolism in cancer. *Trends Biochem. Sci.* 39, 191-198.

156 **van Knegsel, A. T. *et al.*** (2005). Effect of dietary energy source on energy balance, production, metabolic disorders and reproduction in lactating dairy cattle. *Reprod. Nutr. Dev.* 45, 665-688.

157 Nelson, D. L. *et al.* **2008**. Lehninger principles of biochemistry. (Macmillan).

158 **Ingvartsen, K. L. *et al.*** (2003). On the relationship between lactational performance and health: is it yield or metabolic imbalance that cause production diseases in dairy cattle? A position paper. *Livest. Prod. Sci.* 83, 277-308.

159 **Friggens, N. C. *et al.*** (2004). Priming the dairy cow for lactation: a review of dry cow feeding strategies. *Anim. Res.* 53, 453-473.

160 **Bauman, D.** (2000). Regulation of nutrient partitioning during lactation: homeostasis and homeorhesis revisited. *Ruminant physiology: digestion, metabolism, growth and reproduction*, 311-328.

161 **Rezaei, R. *et al.*** (2016). Amino acids and mammary gland development: nutritional implications for milk production and neonatal growth. *J. Anim. Sci. and Biotechnol.* 7, 20.

162 **Boyd, D. R. *et al.*** (1995). Nutrient uptake and endocrine regulation of milk synthesis by mammary tissue of lactating sows. *J. Anim. Sci.* 73, 36-56.

163 **Wehrens, R. *et al.*** (2016). Improved batch correction in untargeted MS-based metabolomics. *Metabolomics* 12, 88.

164 **Moco, S. *et al.*** (2007). Metabolomics technologies and metabolite identification. *TrAC-Trends Anal. Chem.* 26, 855-866.

165 **Zhang, S. *et al.*** (2010). Advances in NMR-based biofluid analysis and metabolite profiling. *Analyst* 135, 1490-1498.

166 **Crockford, D. J. *et al.*** (2006). Statistical heterospectroscopy, an approach to the integrated analysis of NMR and UPLC-MS data sets: application in metabonomic toxicology studies. *Anal. Chem.* 78, 363-371.

167 CVB. **2011**. Chemische samenstellingen en nutritionele waarden van voedermiddelen (in Dutch). (Productschap Diervoeder, CBV (Centraal Veevoeder Bureau)).

168 **CVB**. (2005). Veevoedertabel. Gegevens over chemische samenstelling, verteerbaarheid en voederwaarde van voedermiddelen. CVB, Lelystad, the Netherlands.

169 **Lu, J. *et al.*** (2015). Effect of the DGAT1 K232A genotype of dairy cows on the milk metabolome and proteome. *J. Dairy Sci.* 98, 3460-3469.

170 **Maher, A. D. *et al.*** (2013). Latent biochemical relationships in the blood-milk metabolic axis of dairy cows revealed by statistical integration of 1H NMR spectroscopic data. *J. Proteome Res.* 12, 1428-1435.

171 **Wishart, D. S. *et al.*** (2007). HMDB: the Human Metabolome Database. *Nucleic Acids Res.* 35, D521-526.

172 **Kuhn, N. *et al.*** (1980). Lactose synthesis: the possibilities of regulation. *J. Dairy Sci.* 63, 328-336.

173 **McManaman, J. L. *et al.*** (2003). Mammary physiology and milk secretion. *Adv. Drug Del. Rev.* 55, 629-641.

174 **van Hoeij, R. *et al.*** (2018). Relationship between metabolic status and behavior in dairy cows in week 4 of lactation. *Anim.*, 1-9.

175 **Capuco, A. *et al.*** (2003). Lactation persistency: Insights from mammary cell proliferation studies. *J. Anim. Sci.* 81, 18-31.

176 **Flint, D. J. *et al.*** (1997). Interactions of Prolactin and Growth Hormone (GH) in the Regulation of Mammary Gland Function and Epithelial Cell Survival. *J. Mammary Gland Biol. Neoplasia* 2, 41-48.

177 **Capuco, A. *et al.*** (2001). Mammary cell number, proliferation, and apoptosis during a bovine lactation: relation to milk production and effect of bST1. *J. Dairy Sci.* 84, 2177-2187.

178 **Monks, J. *et al.*** (2008). Epithelial cells remove apoptotic epithelial cells during post-lactation involution of the mouse mammary gland. *Biol. Reprod.* 78, 586-594.

179 **Miller, N. *et al.*** (2006). Effect of stage of lactation and parity on mammary gland cell renewal1. *J. Dairy Sci.* 89, 4669-4677.

180 Akers, R. M. **2016**. Lactation and the mammary gland. (John Wiley & Sons).

181 **Gibellini, F. *et al.*** (2010). The Kennedy pathway—de novo synthesis of phosphatidylethanolamine and phosphatidylcholine. *IUBMB life* 62, 414-428.

182 **Kennedy, E. P. *et al.*** (1956). The function of cytidine coenzymes in the biosynthesis of phospholipides. *J Biol. Chem.* 222, 193-214.

183 **Li, Z. *et al.*** (2008). Thematic review series: glycerolipids. Phosphatidylcholine and choline homeostasis. *J. Lipid Res.* 49, 1187-1194.

184 **O'Riordan, N. *et al.*** (2014). Structural and functional characteristics of bovine milk protein glycosylation. *Glycobiology* 24, 220-236.

185 **Bonfatti, V. *et al.*** (2014). Glycosylation of κ-casein: Genetic and nongenetic variation and effects on rennet coagulation properties of milk. *J. Dairy Sci.* 97, 1961-1969.

186 **Robitaille, G. *et al.*** (1991). Variation in the N-acetyl neuraminic acid content of bovine κ-casein. *J. Dairy Res.* 58, 107-114.

187 **Kates, J. *et al.*** (1967). Poxvirus DNA-dependent RNA polymerase. *Pro. Natl. Acad Sci. U. S. A.* 58, 134.

188 **Iorio, E. *et al.*** (2005). Alterations of choline phospholipid metabolism in ovarian tumor progression. *Cancer Res.* 65, 9369-9376.

189 **Aboagye, E. O. *et al.*** (1999). Malignant transformation alters membrane choline phospholipid metabolism of human mammary epithelial cells. *Cancer Res.* 59, 80-84.

190 **Morrish, F. *et al.*** (2010). Myc-dependent mitochondrial generation of acetyl-CoA contributes to fatty acid biosynthesis and histone acetylation during cell cycle entry. *J. Biol. Chem.* 285, 36267-36274.

191 **Linzell, J. *et al.*** (1976). The secretion of citrate into milk. *J. Physiol.* 260, 739-750.

192 **Baticz, O. *et al.*** (2002). Relationship between concentration of citrate and ketone bodies in cow's milk. *Acta Vet. Hung.* 50, 253-261.

193 **Banks, W. *et al.*** (1984). Effect of inclusion of different forms of dietary fatty acid on the yield and composition of cow's milk. *J. Dairy Res.* 51, 387-395.

194 **Van der Drift, S. *et al.*** (2012). Protein and fat mobilization and associations with serum β-hydroxybutyrate concentrations in dairy cows. *J. Dairy Sci.* 95, 4911-4920.

195 **Whitaker, D. *et al.*** (1993). Some effects of nutrition and management on the fertility of dairy cattle. *Vet. Rec.* 133, 61-64.

196 **Crockford, D. J. *et al.*** (2008). 1H NMR and UPLC-MSE statistical heterospectroscopy: characterization of drug metabolites (xenometabolome) in epidemiological studies. *Anal. Chem.* 80, 6835-6844.

197 **Doepel, L. *et al.*** (2002). Peripartum Performance and Metabolism of Dairy Cows in Response to Prepartum Energy and Protein Intake. *J. Dairy Sci.* 85, 2315-2334.

198 **Raggio, G. *et al.*** (2006). Effect of Casein and Propionate Supply on Mammary Protein Metabolism in Lactating Dairy Cows. *J. Dairy Sci.* 89, 4340-4351.

199 **Wu, G. *et al.*** (2013). Dietary requirements of "nutritionally non-essential amino acids" by animals and humans. *Amino Acids* 44, 1107-1113.

200 **Shibano, K.-i. *et al.*** (2005). The relationship between changes in serum glycine and alanine concentrations in non-essential amino acid and milk production in the transition period in dairy cows. *J. Vet. Med. Sci.* 67, 191-193.

201 **Shields, A. F. *et al.*** (1990). Utilization of labeled thymidine in DNA synthesis: studies for PET. *J. Nucl. Med.* 31, 337-342.

202 **Goff, J. P.** (2008). The monitoring, prevention, and treatment of milk fever and subclinical hypocalcemia in dairy cows. *Vet. J.* 176, 50-57.

203 **Price, P. *et al.*** (1980). New biochemical marker for bone metabolism. Measurement by radioimmunoassay of bone GLA protein in the plasma of normal subjects and patients with bone disease. *J. Clin. Invest.* 66, 878-883.

204 Marshall, W. J. *et al.* **2008**. Clinical biochemistry: metabolic and clinical aspects. (Elsevier Health Sciences).

205 **Kessel, S. *et al.*** (2008). Individual variability in physiological adaptation to metabolic stress during early lactation in dairy cows kept under equal conditions. *J. Dairy Sci.* 86, 2903-2912.

206 **Zhou, Z. *et al.*** (2016). Circulating amino acids in blood plasma during the peripartal period in dairy cows with different liver functionality index. *J. Dairy Sci.* 99, 2257-2267.

207 **Sato, H.** (2009). Increased blood concentration of isopropanol in ketotic dairy cows and isopropanol production from acetone in the rumen. *Animal Sci. J.* 80, 381-386.

208 **LeBlanc, S.** (2010). Monitoring metabolic health of dairy cattle in the transition period. *J. Reprod. Develop.* 56, S29-S35.

209 **Peterson, S. E. *et al.*** (2012). Effects of dietary betaine on milk yield and milk composition of mid-lactation Holstein dairy cows. *J. Dairy Sci.* 95, 6557-6562.

210 **Schwalm, J. W. *et al.*** (1972). Blood Metabolite Interrelationships and Changes in Mammary Gland Metabolism During Subclinical Ketosis1. *J. Dairy Sci.* 55, 58-64.

211 **Carlson, D. *et al.*** (2007). Dietary L-carnitine affects periparturient nutrient metabolism and lactation in multiparous cows. *J. Dairy Sci.* 90, 3422-3441.

212 **Kim, S. W. *et al.*** (2009). Regulatory role for amino acids in mammary gland growth and milk synthesis. *Amino Acids* 37, 89-95.

213 **Stentoft, C. *et al.*** (2015). Absorption and intermediary metabolism of purines and pyrimidines in lactating dairy cows. *Br. J. Nutr.* 113, 560-573.

214 **Giesecke, D. *et al.*** (1994). Mammary and Renal Excretion of Purine Metabolites in Relation to Energy Intake and Milk Yield in Dairy Cows. *J. Dairy Sci.* 77, 2376-2381.

215 **Hammon, H. *et al.*** (2009). Performance and metabolic and endocrine changes with emphasis on glucose metabolism in high-yielding dairy cows with high and low fat content in liver after calving. *J. Dairy Sci.* 92, 1554-1566.

216 **Gross, J. *et al.*** (2011). Performance and metabolic profile of dairy cows during a lactational and deliberately induced negative energy balance with subsequent realimentation. *J. Dairy Sci.* 94, 1820-1830.

217 **Bauman, D. E. *et al.*** (1980). Partitioning of nutrients during pregnancy and lactation: a review of mechanisms involving homeostasis and homeorhesis. *J. Dairy Sci.* 63, 1514-1529.

218 **Kleiber, M. *et al.*** (1955). Glucose as a precursor of milk constituents in the intact dairy cow. *Biochim. Biophys. Acta* 17, 252-260.

219 **Tedeschi, P. M. *et al.*** (2013). Contribution of serine, folate and glycine metabolism to the ATP, NADPH and purine requirements of cancer cells. *Cell Death Dis.* 4, e877.

220 **Laarveld, B. *et al.*** (1981). The effect of insulin on net metabolism of glucose and amino acids by the bovine mammary gland. *Endocrinology* 108, 2217-2221.

221 **Drackley, J. K. *et al.*** (2001). Adaptations of glucose and long-chain fatty acid metabolism in liver of dairy cows during the periparturient period. *J. Dairy Sci.* 84, E100-E112.

222 **Li, L. O. *et al.*** (2015). Compartmentalized acyl-CoA metabolism in skeletal muscle regulates systemic glucose homeostasis. *Diabetes* 64, 23-35.

223 **Komaragiri, M. V. *et al.*** (1998). Factors affecting body tissue mobilization in early lactation dairy cows. 2. Effect of dietary fat on mobilization of body fat and protein. *J. Dairy Sci.* 81, 169-175.

224 **Corbin, K. D. *et al.*** (2012). Choline metabolism provides novel insights into nonalcoholic fatty liver disease and its progression. *Curr. Opin. Gastroenterol.* 28, 159-165.

225 **Castillo, C. *et al.*** (2006). Plasma malonaldehyde (MDA) and total antioxidant status (TAS) during lactation in dairy cows. *Res. Vet. Sci.* 80, 133-139.

226 **Brockman, R. P. *et al.*** (1986). Hormonal regulation of metabolism in ruminants; a review. *Livest. Prod. Sci.* 14, 313-334.

227 **Fröhli, D. M. *et al.*** (1988). Nonesterified Fatty Acids and Glucose in Lactating Dairy Cows: Diurnal Variations and Changes in Responsiveness During Fasting to Epinephrine and Effects of Beta-Adrenergic Blockade. *J. Dairy Sci.* 71, 1170-1177.

228 **Whitton, P. D. *et al.*** (1978). Stimulation by vasopressin, angiotensin and oxytocin of gluconeogenesis in hepatocyte suspensions. *Biochem. J.* 176, 893-898.

229 **Udenfriend, S. *et al.*** (1956). Precursors of adrenal epinephrine and norepinephrine in vivo. *Biochim. Biophys. Acta* 20, 48-52.

230 **Davis, S. R.** ***et al.*** (2004). Milk L-lactate concentration is increased during mastitis. *J. Dairy Res.* 71, 175-181.

231 **Zahra, L. C.** ***et al.*** (2006). Effects of Rumen-Protected Choline and Monensin on Milk Production and Metabolism of Periparturient Dairy Cows. *J. Dairy Sci.* 89, 4808-4818.

232 **Pinotti, L.** ***et al.*** (2003). Rumen-protected choline administration to transition cows: effects on milk production and vitamin E status. *J. Vet. Med. A* 50, 18-21.

233 **Piepenbrink, M.** ***et al.*** (2003). Liver Metabolism and Production of Cows Fed Increasing Amounts of Rumen-Protected Choline During the Periparturient Period1. *J. Dairy Sci.* 86, 1722-1733.

234 **Guretzky, N. A. J.** ***et al.*** (2006). Lipid Metabolite Profiles and Milk Production for Holstein and Jersey Cows Fed Rumen-Protected Choline During the Periparturient Period1. *J. Dairy Sci.* 89, 188-200.

235 **Buitenhuis, A. J.** ***et al.*** (2013). Estimation of genetic parameters and detection of quantitative trait loci for metabolites in Danish Holstein milk. *J. Dairy Sci.* 96, 3285-3295.

236 **Tacoma, R.** ***et al.*** (2016). Characterization of the bovine milk proteome in early-lactation Holstein and Jersey breeds of dairy cows. *J. Proteomics* 130, 200-210.

237 **D'Alessandro, A.** ***et al.*** (2011). The bovine milk proteome: cherishing, nourishing and fostering molecular complexity. An interactomics and functional overview. *Mol. BioSyst.* 7, 579-597.

238 **Zhang, L.** ***et al.*** (2015). Bovine milk proteome in the first 9 days: protein interactions in maturation of the immune and digestive system of the newborn. *PloS one* 10, e0116710.

239 **Aletta, J. M.** ***et al.*** (1998). Protein methylation: a signal event in post-translational modification. *Trends Biochem. Sci.* 23, 89-91.

240 **Comb, D. G.** ***et al.*** (1966). The methylation of lysine residues in protein. *J. Biol. Chem.* 241, 1857-1862.

241 **Lu, J.** ***et al.*** (2011). Filter-aided sample preparation with dimethyl labeling to identify and quantify milk fat globule membrane proteins. *J. roteomics* 75, 34-43.

242 **Cox, J.** ***et al.*** (2008). MaxQuant enables high peptide identification rates, individualized ppb-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* 26, 1367.

243 **Bedford, M. T.** (2007). Arginine methylation at a glance. *J. Cell Sci.* 120, 4243.

244 **Boisvert, F.-M.** ***et al.*** (2003). A proteomic analysis of arginine-methylated protein complexes. *Mol. Cell. Proteomics* 2, 1319-1330.

245 **Bedford, M. T.** ***et al.*** (2005). Arginine methylation: an emerging regulatorof protein function. *Mol. Cell* 18, 263-272.

246 **Kamen, B. A.** ***et al.*** (2004). A review of folate receptor alpha cycling and 5-methyltetrahydrofolate accumulation with an emphasis on cell models in vitro. *Adv. Drug Del. Rev.* 56, 1085-1097.

247 **Sandoval, R. M.** ***et al.*** (2004). Uptake and trafficking of fluorescent conjugates of folic acid in intact kidney determined using intravital two-photon microscopy. *Am. J. Physiol. Cell Physiol.* 287, C517-C526.

248 **Ingvartsen, K. L.** ***et al.*** (2000). Integration of metabolism and intake regulation: a review focusing on periparturient animals. *J. Dairy Sci.* 83, 1573-1597.

249 **Drackley, J. K.** (1999). Biology of dairy cows during the transition period: The final frontier? *J. Dairy Sci.* 82, 2259-2273.

250 **Butler, W. R.** (2003). Energy balance relationships with follicular development, ovulation and fertility in postpartum dairy cows. *Livest. Prod. Sci.* 83, 211-218.

251 **Pezeshki, A. *et al.*** (2007). Effects of short dry periods on performance and metabolic status in Holstein dairy cows. *J. Dairy Sci.* 90, 5531-5541.

252 **Dann, H. *et al.*** (2005). Prepartum intake, postpartum induction of ketosis, and periparturient disorders affect the metabolic status of dairy cows. *J. Dairy Sci.* 88, 3249-3264.

253 **De Feu, M. *et al.*** (2009). The effect of dry period duration and dietary energy density on milk production, bioenergetic status, and postpartum ovarian function in Holstein-Friesian dairy cows. *J. Dairy Sci.* 92, 6011-6022.

254 **McArt, J. A. *et al.*** (2013). Elevated non-esterified fatty acids and β-hydroxybutyrate and their association with transition dairy cow performance. *Vet. J.* 198, 560-570.

255 **Hostens, M. *et al.*** (2012). On-farm evaluation of the effect of metabolic diseases on the shape of the lactation curve in dairy cows through the MilkBot lactation model. *J. Dairy Sci.* 95, 2988-3007.

256 **Nielsen, N. I. *et al.*** (2005). Predicting risk of ketosis in dairy cows using in-line measurements of β-hydroxybutyrate: A biological model. *J. Dairy Sci.* 88, 2441-2453.

257 **van Knegsel, A. *et al.*** (2010). Ketone body concentration in milk determined by Fourier transform infrared spectroscopy: Value for the detection of hyperketonemia in dairy cows. *J. Dairy Sci.* 93, 3065-3069.

258 **Hall, M. A.** (1999). Correlation-based feature selection for machine learning.

259 **Liechty, J. C. *et al.*** (2004). Bayesian correlation estimation. *Biometrika* 91, 1-14.

260 **Tremblay, M. *et al.*** (2018). Identifying poor metabolic adaptation during early lactation in dairy cows using cluster analysis. *J. Dairy Sci.* 101, 7311-7321.

261 **Vicari, T. *et al.*** (2008). Postprandial blood hormone and metabolite concentrations influenced by feeding frequency and feeding level in veal calves. *Domest. Anim. Endocrinol.* 34, 74-88.

262 **Graber, M. *et al.*** (2012). Identification of plasma and hepatic parameters related to metabolic robustness in dairy cows. *J. Anim. Physiol. Anim. Nutr.* 96, 75-84.

263 Borcard, D. *et al.* **2018**. Numerical ecology with R. (Springer).

264 Johnson, R. A. *et al.* **2002**. Applied multivariate statistical analysis. Vol. 5 (Prentice hall Upper Saddle River, NJ).

265 **Rousseeuw, P. J.** (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* 20, 53-65.

266 **Wong, H. B. *et al.*** (2011). Measures of diagnostic accuracy: sensitivity, specificity, PPV and NPV. *Proceedings of Singapore healthcare* 20, 316-318.

267 Louppe, G. *et al.* in *27th Annual Conference on Neural Information Processing Systems* 431-439 (Neural Information Processing Systems (NIPS).).

268 **Breiman, L.** (2001). Random forests. *Mach. Learn.* 45, 5-32.

269 **Francisco, C. *et al.*** (2002). Propionibacteria Fed to Dairy Cows: Effects on Energy Balance, Plasma Metabolites and Hormones, and Reproduction1. *J. Dairy Sci.* 85, 1738-1751.

270 **Van Knegsel, A. *et al.*** (2007). Dietary energy source in dairy cows in early lactation: metabolites and metabolic hormones. *J. Dairy Sci.* 90, 1477-1485.

271 Hair, J. F. **2006**. Multivariate data analysis. (Pearson Education India).

272 **Franke, G. R.** (2010). Multicollinearity. *Wiley International Encyclopedia of Marketing*.

273 **Borcard, D. *et al.*** (2011). Numerical ecologywith R. *Springer, New YorkBouchard M, Pothier D, Gauthier S (2008) Fire return intervals and tree species succession in the North Shore region of eastern Quebec. Can. J. For Res.* 38, 16211633.

274 Duffield, T. *et al.* in *Southwest Nutrition and Management Conference.* 106-114.

275 **Lomander, H. *et al.*** (2012). Supplemental feeding with glycerol or propylene glycol of dairy cows in early lactation—Effects on metabolic status, body condition, and milk yield. *J. Dairy Sci.* 95, 2397-2408.

276 **Bauman, D. *et al.*** (1983). Control of nutrient partitioning in lactating ruminants. *Biochemistry of lactation* 14, 437-462.

277 **Lucy, M. *et al.*** (1992). Influence of diet composition, dry-matter intake, milk production and energy balance on time of post-partum ovulation and fertility in dairy cows. *Anim. Sci.* 54, 323-331.

278 **McGuire, M. A. *et al.*** (1995). Insulin regulates circulating insulin-like growth factors and some of their binding proteins in lactating cows. *Am. J. Physiol. Endocrinol. Metabol.* 269, E723-E730.

279 **Harding, J. *et al.*** (1994). Insulin-like growth factor 1 alters feto-placental protein and carbohydrate metabolism in fetal sheep. *Endocrinology* 134, 1509-1514.

280 **Rigout, S. *et al.*** (2002). Duodenal glucose increases glucose fluxes and lactose synthesis in grass silage-fed dairy cows. *J. Dairy Sci.* 85, 595-606.

281 **Gems, D. *et al.*** (2001). Insulin/IGF signalling and ageing: seeing the bigger picture. *Curr. Opin. Genet. Dev.* 11, 287-292.

282 **Sonntag, W. E. *et al.*** (2005). Growth hormone and insulin-like growth factor-1 (IGF-1) and their influence on cognitive aging. *Ageing Res. Rev.* 4, 195-212.

283 **Rutten, C. *et al.*** (2013). Invited review: Sensors to support health management on dairy farms. *J. Dairy Sci.* 96, 1928-1952.

284 **Dewhurst, R. *et al.*** (2000). Effects of altering energy and protein supply to dairy cows during the dry period. 1. Intake, body condition, and milk production. *J. Dairy Sci.* 83, 1782-1794.

285 **Caraviello, D. *et al.*** (2006). Analysis of reproductive performance of lactating cows on large dairy farms using machine learning algorithms. *J. Dairy Sci.* 89, 4703-4722.

286 **Grzesiak, W. *et al.*** (2006). Methods of predicting milk yield in dairy cows—Predictive capabilities of Wood's lactation curve and artificial neural networks (ANNs). *Comput. Electron. Agric.* 54, 69-83.

287 **West, J. *et al.*** (2003). Effects of hot, humid weather on milk temperature, dry matter intake, and milk yield of lactating dairy cows. *J. Dairy Sci.* 86, 232-242.

288 **Pralle, R. *et al.*** (2018). Predicting blood β-hydroxybutyrate using milk Fourier transform infrared spectrum, milk composition, and producer-reported variables

with multiple linear regression, partial least squares regression, and artificial neural network. *J. Dairy Sci.* 101, 4378-4387.

289 **Duffield, T. *et al.*** (2009). Impact of hyperketonemia in early lactation dairy cows on health and production. *J. Dairy Sci.* 92, 571-580.

290 **McArt, J. *et al.*** (2013). Dry period and parturient predictors of early lactation hyperketonemia in dairy cattle. *J. Dairy Sci.* 96, 198-209.

291 **Geishauser, T. *et al.*** (1997). An evaluation of milk ketone tests for the prediction of left displaced abomasum in dairy cows. *J. Dairy Sci.* 80, 3188-3192.

292 **Ehret, A. *et al.*** (2015). Use of genomic and metabolic information as well as milk performance records for prediction of subclinical ketosis risk via artificial neural networks. *J. Dairy Sci.* 98, 322-329.

293 **Lean, I. *et al.*** (1992). Time series cross-correlation analysis of postparturient relationships among serum metabolites and yield variables in Holstein cows. *J. Dairy Sci.* 75, 1891-1900.

294 **van Knegsel, A. *et al.*** (2007). Effect of glucogenic vs. lipogenic diets on energy balance, blood metabolites, and reproduction in primiparous and multiparous dairy cows in early lactation. *J. Dairy Sci.* 90, 3397-3409.

295 **Grelet, C. *et al.*** (2016). Development of Fourier transform mid-infrared calibrations to predict acetone, β-hydroxybutyrate, and citrate contents in bovine milk through a European dairy network. *J. Dairy Sci.* 99, 4816-4825.

296 **Dohoo, I. R. *et al.*** (1984). Subclinical ketosis: prevalence and associations with production and disease. *Can. J. Comp. Med.* 48, 1.

297 **Oikonomou, G. *et al.*** (2008). Genetic profile of body energy and blood metabolic traits across lactation in primiparous Holstein cows. *J. Dairy Sci.* 91, 2814-2822.

298 **Halachmi, I. *et al.*** (2004). Predicting Feed Intake of the Individual Dairy Cow. *J. Dairy Sci.* 87, 2254-2267.

299 **Staples, C. R. *et al.*** (1990). Relationship Between Ovarian Activity and Energy Status During the Early Postpartum Period of High Producing Dairy Cows1, 2. *J. Dairy Sci.* 73, 938-947.

300 **Hurtaud, C. *et al.*** (2000). Effect of graded duodenal infusions of glucose on yield and composition of milk from dairy cows. 2. Diets based on grass silage. *J. Dairy Sci.* 83, 2952-2962.

301 **Lucy, M.** (2001). Reproductive loss in high-producing dairy cattle: where will it end? *J. Dairy Sci.* 84, 1277-1293.

302 **Ingvartsen, K. L.** (2006). Feeding-and management-related diseases in the transition cow: Physiological adaptations around calving and strategies to reduce feeding-related diseases. *Anim. Feed Sci. Technol.* 126, 175-213.

303 **Vinayavekhin, N. *et al.*** (2009). Exploring disease through metabolomics. *ACS Chem. Biol.* 5, 91-103.

304 **Smolinska, A. *et al.*** (2012). NMR and pattern recognition methods in metabolomics: from data acquisition to biomarker discovery: a review. *Anal. Chim. Acta* 750, 82-97.

305 **Ilves, A. *et al.*** (2012). Alterations in milk and blood metabolomes during the first months of lactation in dairy cows. *J. Dairy Sci.* 95, 5788-5797.

306 **Hailemariam, D. *et al.*** (2014). Metabolomics Approach Reveals Altered Plasma Amino Acid and Sphingolipid Profiles Associated with Patholological State in Transition Dairy Cows. *Curr. Metabolomics* 2, 184-195.

307 **Tian, H. *et al.*** (2015). Identification of diagnostic biomarkers and metabolic pathway shifts of heat-stressed lactating dairy cows. *J. Proteomics* 125, 17-28.

308 **Boudonck, K. J. *et al.*** (2009). Characterization of the biochemical variability of bovine milk using metabolomics. *Metabolomics* 5, 375-386.

309 **Kohl, S. M. *et al.*** (2012). State-of-the art data normalization methods improve NMR-based metabolomic analysis. *Metabolomics* 8, 146-160.

310 **Sun, H.-Z. *et al.*** (2015). Metabolomics of four biofluids from dairy cows: potential biomarkers for milk production and quality. *J. Proteome Res.* 14, 1287-1298.

311 **Bertram, H. C. *et al.*** (2011). Effect of dietary nitrogen content on the urine metabolite profile of dairy cows assessed by nuclear magnetic resonance (NMR)-based metabolomics. *J. Agric. Food Chem.* 59, 12499-12505.

312 **Ametaj, B. N. *et al.*** (2010). Metabolomics reveals unhealthy alterations in rumen metabolism with increased proportion of cereal grain in the diet of dairy cows. *Metabolomics* 6, 583-594.

313 **Libbrecht, M. W. *et al.*** (2015). Machine learning applications in genetics and genomics. *Nat. Rev. Genet.* 16, 321.

314 **Taylor, R.** (1990). Interpretation of the correlation coefficient: a basic review. *J. Diagn. Med. Sonogr.* 6, 35-39.

315 **Ammer, S. *et al.*** (2018). Impact of diet composition and temperature–humidity index on water and dry matter intake of high-yielding dairy cows. *J. Anim. Physiol. Anim. Nutr.* 102, 103-113.

316 **De Marchi, M. *et al.*** (2014). Invited review: Mid-infrared spectroscopy as phenotyping tool for milk traits1. *J. Dairy Sci.* 97, 1171-1186.

317 **Melzer, N. *et al.*** (2013). Investigating associations between milk metabolite profiles and milk traits of Holstein cows. *J. Dairy Sci.* 96, 1521-1534.

318 **Edmonson, A. J. *et al.*** (1989). A Body Condition Scoring Chart for Holstein Dairy Cows. *J. Dairy Sci.* 72, 68-78.

319 **Pires, J. A. *et al.*** (2013). Effects of body condition score at calving on indicators of fat and protein mobilization of periparturient Holstein-Friesian cows. *J. Dairy Sci.* 96, 6423-6439.

320 **Megahed, A. A. *et al.*** (2019). Changes in skeletal muscle thickness and echogenicity and plasma creatinine concentration as indicators of protein and intramuscular fat mobilization in periparturient dairy cows. *J. Dairy Sci.* 102, 5550-5565.

321 **Butler, W. *et al.*** (1981). The relationships between energy balance, milk production and ovulation in postpartum Holstein cows. *J. Anim. Sci.* 53, 742-748.

322 **Bouraoui, R. *et al.*** (2002). The relationship of temperature-humidity index with milk production of dairy cows in a Mediterranean climate. *Anim. Res.* 51, 479-491.

323 **Kelly, J.** (1977). Changes in serum B hydroxybutyrate concentrations in dairy cows kept under commercial farm conditions. *Vet. Rec.* 101, 499-502.

324 **Andersen, J. B. *et al.*** (2005). The Effects of Dry Period Versus Continuous Lactation on Metabolic Status and Performance in Periparturient Cows. *J. Dairy Sci.* 88, 3530-3541.

325 **Sundekilde, U. K. *et al.*** (2013). NMR-Based Milk Metabolomics. *Metabolites* 3, 204-222.

326 **DERVISHI, E. *et al.*** (2016). GC-MS metabolomics identifies metabolite alterations that precede subclinical mastitis in the blood of transition dairy cows. *J. Proteome Res.*

327 **Imhasly, S. *et al.*** (2014). Metabolomic biomarkers correlating with hepatic lipidosis in dairy cows. *BMC Vet. Res.* 10, 122.

328 **BAŞOĞLU, A. *et al.*** (2014). NMR-based metabolomic evaluation in dairy cows with displaced abomasum. *Turk. J. Vet. and Anim. Sci.* 38, 325-330.

329 **Ospina, P. *et al.*** (2010). Associations of elevated nonesterified fatty acids and β-hydroxybutyrate concentrations with early lactation reproductive performance and milk production in transition dairy cattle in the northeastern United States. *J. Dairy Sci.* 93, 1596-1603.

330 Brockman, R. in *Quantitative aspects of ruminant digestion and metabolism.* 291-310 (CAB International, Wallingford, UK, 2005).

331 **Danfær, A.** (1994). Nutrient metabolism and utilization in the liver. *Livest. Prod. Sci.* 39, 115-127.

332 **Wiltrout, D. W. *et al.*** (1972). Contribution of Propionate to Glucose Synthesis in the Lactating and Nonlactating Cow1. *J. Dairy Sci.* 55, 307-317.

333 **Bauman, D. *et al.*** (1971). Propionate production in the rumen of cows fed either a control or high-grain, low-fiber diet. *J. Dairy Sci.* 54, 1282-1287.

334 **Russell, J.** (1998). The importance of pH in the regulation of ruminal acetate to propionate ratio and methane production in vitro. *J. Dairy Sci.* 81, 3222-3230.

335 Stangassinger, M. *et al.* in *Proc. Soc. Nutr. Physiol.* 162-171.

336 **Miyoshi, S. *et al.*** (2001). Effects of propylene glycol drenching on energy balance, plasma glucose, plasma insulin, ovarian function and conception in dairy cows. *Anim. Reprod. Sci.* 68, 29-43.

337 **Komatsu, T. *et al.*** (2005). Changes in gene expression of glucose transporters in lactating and nonlactating cows. *J. Anim. Sci.* 83, 557-564.

338 **Bonczek, R. *et al.*** (1988). Responses of somatotropin, insulin, prolactin, and thyroxine to selection for milk yield in Holsteins. *J. Dairy Sci.* 71, 2470-2479.

339 **Bichell, D. P. *et al.*** (1992). Growth hormone rapidly activates insulin-like growth factor I gene transcription in vivo. *Mol. Endocrinol.* 6, 1899-1908.

340 **Le Roith, D. *et al.*** (2001). The somatomedin hypothesis: 2001. *Endocr. Rev.* 22, 53-74.

341 **Thissen, J.-P. *et al.*** (1994). Nutritional regulation of the insulin-like growth factors. *Endocr. Rev.* 15, 80-101.

342 **Vicini, J. L. *et al.*** (1991). Nutrient balance and stage of lactation affect responses of insulin, insulin-like growth factors I and II, and insulin-like growth factor-binding protein 2 to somatotropin administration in dairy cows. *J. Nutr.* 121, 1656-1664.

343 **Sharma, B. *et al.*** (1994). Expression of insulin-like growth factor-I in cows at different stages of lactation and in late lactation cows treated with somatotropin. *J. Dairy Sci.* 77, 2232-2241.

344 **Vandehaar, M. *et al.*** (1995). Effect of dietary energy restriction on the expression of insulin-like growth factor-I in liver and corpus luteum of heifers. *J. Dairy Sci.* 78, 832-841.

345 **Lindmark Månsson, H.** (2008). Fatty acids in bovine milk fat. *Food Nutr. Res.* 52, 1821.

346 **Nelson, D.** ***et al.*** (2008). Lehninger principles of biochemistry, WH Freeman and Company. *New York, NY, USA*.

347 **Konar, A.** ***et al.*** (1971). The concentrations of some water-soluble constituents in the milks of cows, sows, ewes and goats. *J. Dairy Res.* 38, 333-341.

348 **Holt, C.** ***et al.*** (1979). Inorganic constituents of milk: I. Correlation of soluble calcium with citrate in bovine milk. *J. Dairy Res.* 46, 433-439.

349 Jeremy M. Berg, L. S., John Tymoczko, Gregory in *Enlarge Biochemistry (9th Edition)* (WH Freeman, 2019).

350 **Podo, F.** (1999). Tumour phospholipid metabolism. *NMR Biomed.* 12, 413-439.

351 **Mann, M.** ***et al.*** (2003). Proteomic analysis of post-translational modifications. *Nat. Biotechnol.* 21, 255.

352 **Zom, R.** ***et al.*** (2011). Effect of rumen-protected choline on performance, blood metabolites, and hepatic triacylglycerols of periparturient dairy cattle. *J. Dairy Sci.* 94, 4016-4027.

353 **Shahsavari, A.** ***et al.*** (2016). The role of rumen-protected choline in hepatic function and performance of transition dairy cows. *Br. J. Nutr.* 116, 35-44.

Appendix

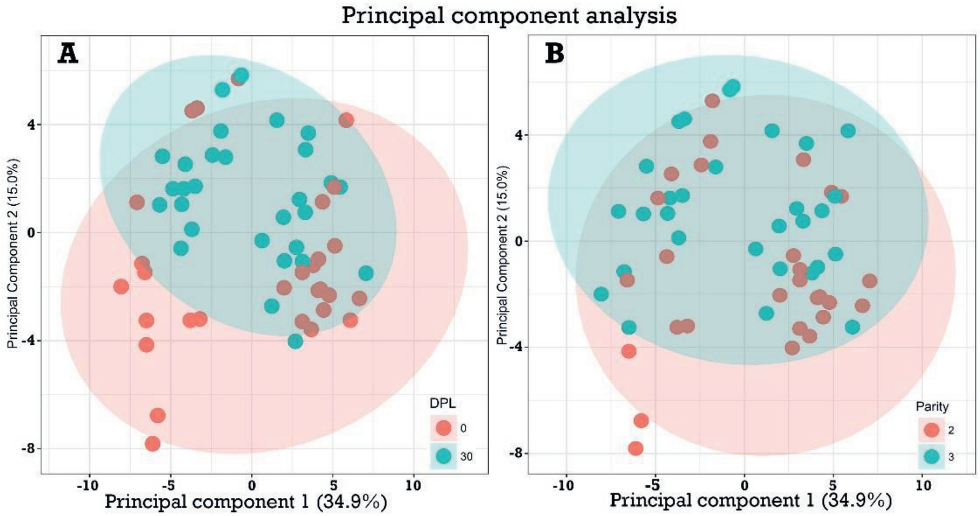


Fig. 2.1. Principal component analyses to identify metabotypes, data were labelled with different dry period length (a) and parity (b).

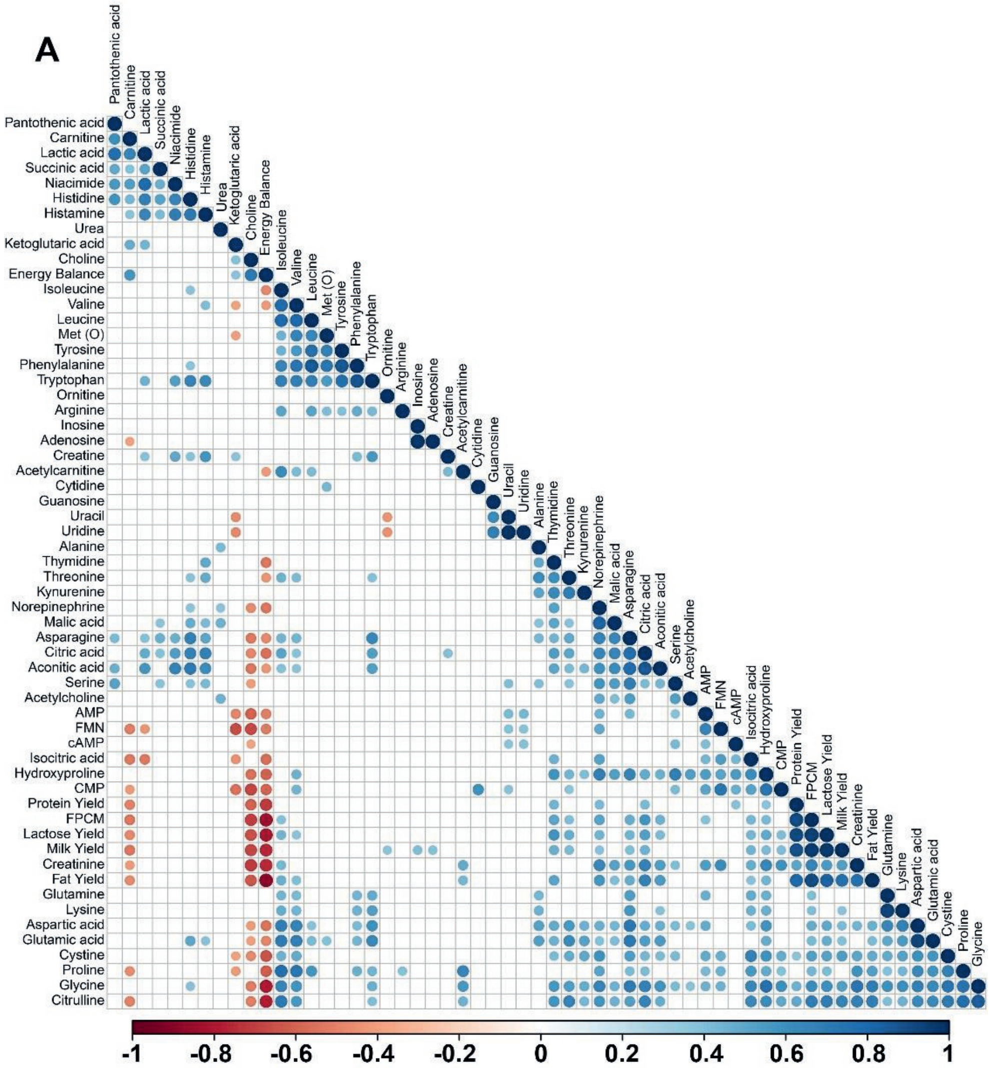


Fig. 2.2. Pearson correlations matrix among milk metabolites, milk production traits and energy balance of dairy cows in week 2 (A). Blank represents the *P*-value of correlation between two variables is insignificant ($P > 0.05$).

For a better digital version, please check online (https://www.nature.com/articles/s41598-018-34190-4#Sec18).

Abbreviations: cAMP, adenosine 3,5-cyclic monophosphate; AMP, adenosine monophosphate; CMP, cytidine monophosphate; FMN, flavin mononucleotide; FPCM, fat- and protein-corrected milk production; Methionine (O), methionine sulfoxide.

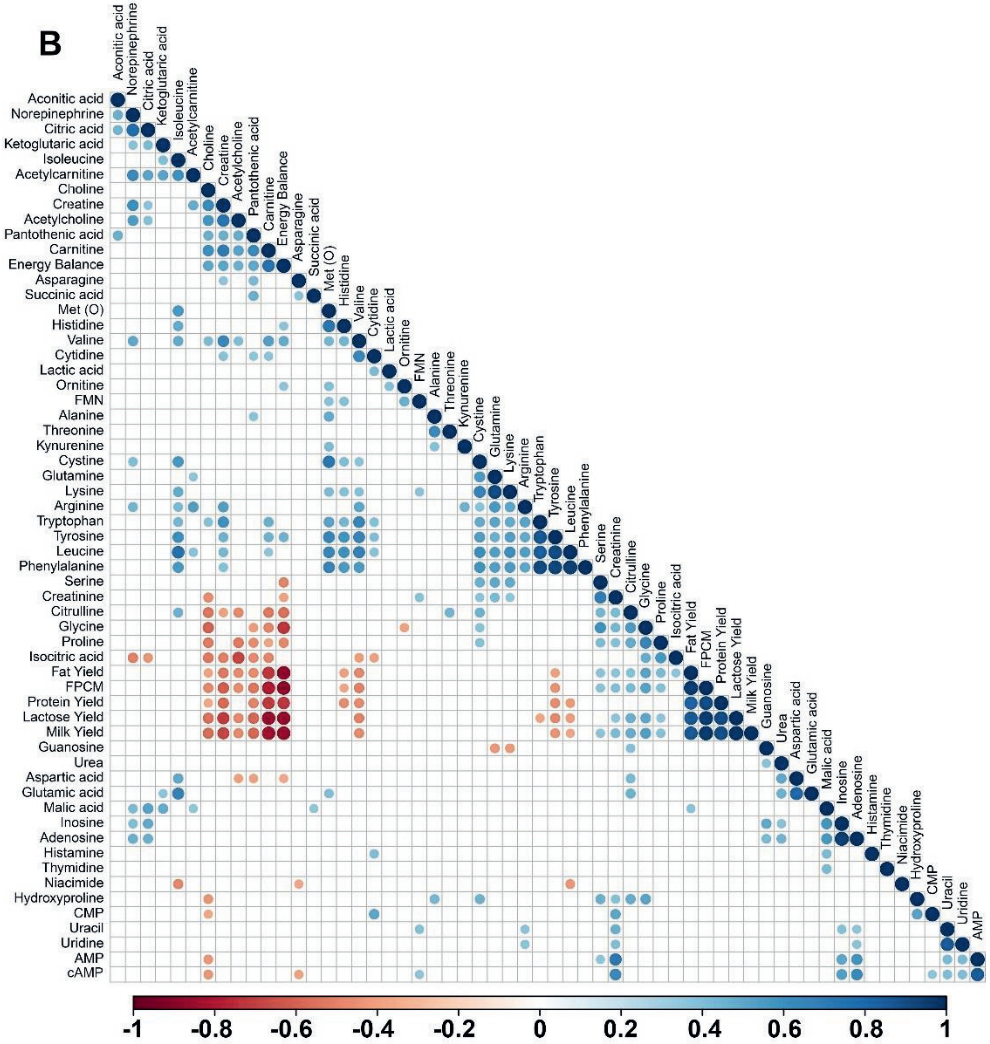


Fig. 2.2. Pearson correlations matrix among milk metabolites, milk production traits and energy balance of dairy cows in week 7 (B). Blank represents the *P*-value of correlation between two variables is insignificant ($P > 0.05$).

For a better digital version, please check online (https://www.nature.com/articles/s41598-018-34190-4#Sec18).

Abbreviations: cAMP, adenosine 3,5-cyclic monophosphate; AMP, adenosine monophosphate; CMP, cytidine monophosphate; FMN, flavin mononucleotide; FPCM, fat- and protein-corrected milk production; Methionine (O), methionine sulfoxide.

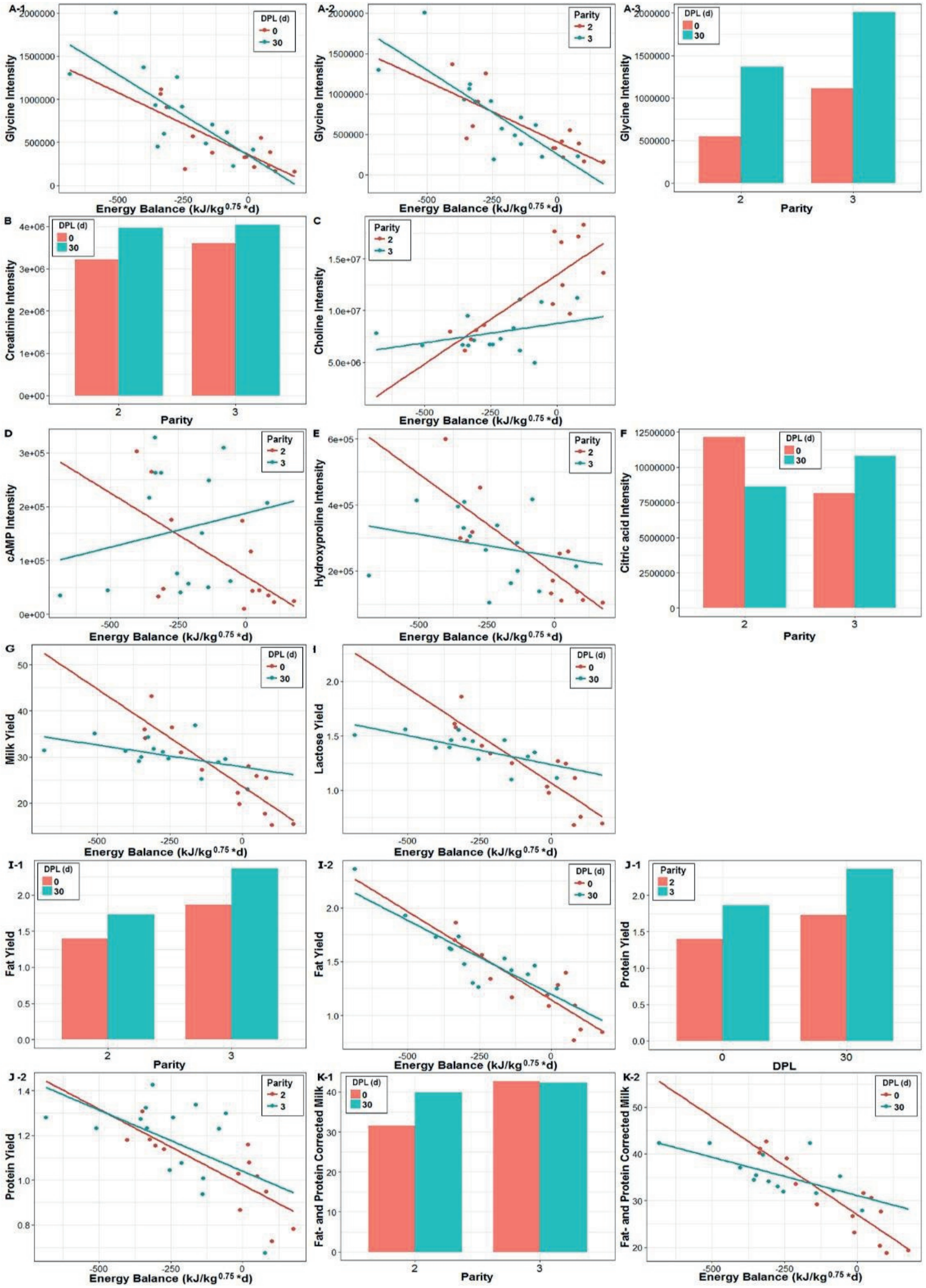


Fig. 2.3. The effect of significant two-way interaction between energy balance (EB) and different treatments, including dry period length (DPL) and parity in week 2 glycine (A), creatinine (B), choline (C), adenosine 3,5-cyclic monophosphate (cAMP) (D), hydroxyproline (E), citric acid (F), milk yield (G), lactose (H), fat yield (I), protein (J), and fat- and protein-corrected milk yield (K).

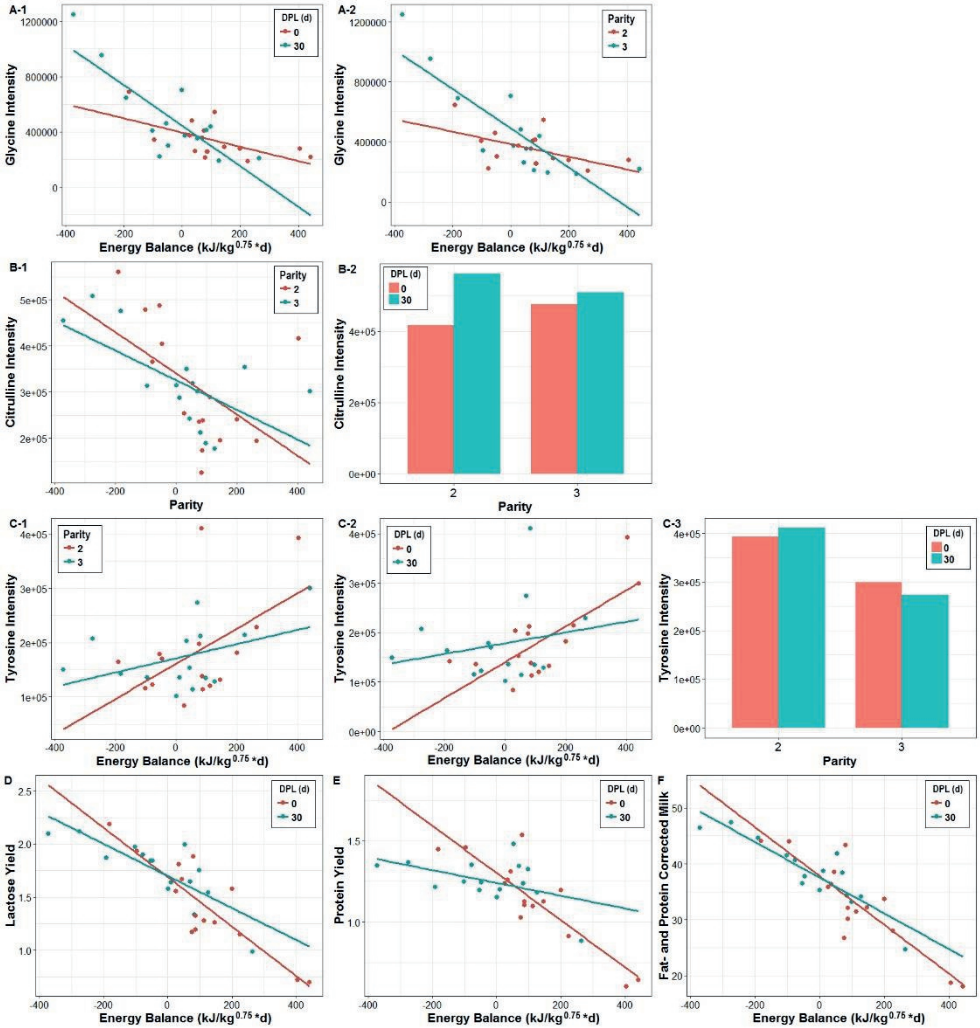


Fig. 2.4. The effect of significant two-way interaction between EB and different treatments, including dry period length (DPL) and parity in week 7 on glycine (A), citrulline (B), tyrosine (C), lactose (D), protein (E), and fat- and protein-corrected milk yield (F).

Table 3.1. The milk metabolites detected either by Liquid chromatography-mass spectrometry (LC-MS), or by nuclear magnetic resonance (NMR), or by both.

No.	**Milk**	**Function**	**Detection by**		**Integrated region in NMR**
1	Acetyl-carnitine	Lipid metabolism	LC-MS	NMR	2.149-2.146; 3.200-3.189
2	Acetate	Energy metabolism		NMR	1.931_1.919
3	Acetone	Lipid Metabolism		NMR	2.243-2.234
4	Acetylcholine	Energy metabolism	LC-MS		
5	Aconitate	Energy metabolism	LC-MS		
6	cis-Aconitate	Energy metabolism		NMR	5.729_5.709
7	Adenosine	Nucleic metabolism	LC-MS		
8	Alanine	AAs metabolism		NMR	1.500-1.474
9	Arginine	AAs metabolism/Urea Cycle	LC-MS		
10	Asparagine	AAs metabolism	LC-MS		
11	Aspartate	AAs metabolism	LC-MS		
12	BHB	Lipid metabolism		NMR	1.203-1.194; 1.192-1.186
13	Butyrate	Lipid metabolism		NMR	2.167-2.160; 2.154-2.150; 1.588-1.538
14	Carnitine	Lipid metabolism	LC-MS		
15	cAMP	Nucleotide metabolism	LC-MS		
16	Choline	Lipid metabolism	LC-MS	NMR	3.210-3.200
17	Citrate	Energy/Lipid metabolism		NMR	2.722-2.641; 2.581-2.508
18	Citrulline	AAs metabolism/Urea cycle	LC-MS		
19	Creatine	Protein degradation		NMR	3.044-3.034
20	Creatinine	Protein degradation		NMR	3.049-3.044
21	CMP	Nucleotide metabolism	LC-MS	NMR	7.986-7.957; 6.007-5.983
22	Cytidine	Nucleotide metabolism	LC-MS		
23	Cystine	AAs metabolism	LC-MS		
24	Ethanolamine	Lipid metabolism		NMR	3.145-3.139
25	FMN	Energy metabolism	LC-MS		
26	Fucose [1]	Lipid metabolism		NMR	1.250-1.242
27	Formate	Energy metabolism		NMR	8.467-8.452
28	Fumarate	Energy metabolism		NMR	6.538-6.517
29	Galactose-1-P	Lactose synthesis		NMR	5.525-5.486
30	Glucose-1-P	Lactose synthesis		NMR	5.475-5.447
31	Glutamate	AAs metabolism	LC-MS	NMR	2.145-2.142; 2.376-2.338
32	Glutamine	AAs metabolism	LC-MS		
33	Glycine	AAs metabolism	LC-MS		
34	GPC	Lipid metabolism		NMR	3.242-3.227
35	Guanosine	Nucleotide metabolism	LC-MS		
36	Hippurate	AAs/Energy metabolism		NMR	7.860-7.822; 7.666-7.623; 7.585-7.533
37	Histidine	AAs metabolism	LC-MS		
38	Hydroxyproline	AAs metabolism	LC-MS		
39	α-Ketoglutarate	Energy metabolism	LC-MS	NMR	3.027-2.999
40	Kynurenine	AAs metabolism	LC-MS		
41	Lactose	Lactose synthesis		NMR	5.263-5.219; 4.507-4.415; 3.185-3.171

Table 3.1. (continued)

42	Lactate	Energy Metabolism		NMR	1.346-1.321
43	Leucine	AAs metabolism	LC-MS		
44	Lysine	AAs metabolism	LC-MS		
45	Methionine [1]	AAs metabolism	LC-MS		
46	Met (O)	AAs metabolism	LC-MS		
47	Nac-Gal/Nac-Glu	Lipid metabolism		NMR	5.283-5.266; 5.219-5.202; 2.067-2.054
48	Nac-NA	Lipd-metabolism		NMR	2.049-2.039
49	Isoleucine	AAs metabolism	LC-MS		
50	Orotate	Nucleotide metabolism		NMR	6.208-6.189
51	Pantothenate	Energy/Lipid metabolism	LC-MS		
52	Phosphocholine	Lipid metabolism		NMR	3.227-3.213
53	Phenylalanine	AAs metabolism	LC-MS		
54	Phosphocreatine	AAs metabolism		NMR	3.055-3.050
55	Proline	AAs metabolism	LC-MS		
56	Serine	AAs metabolism	LC-MS		
57	TMAO	AAs metabolism		NMR	3.273-3.263
58	UDP-Nac-Gal	Lipid metabolism		NMR	8.084-8.040; 5.417-5.389; 2.082-2.074
59	UDP-Nac-Glu	Lipid metabolism		NMR	8.168-8.099; 5.376-5.344; 2.074-2.066
60	Uracil	Nucleotide metabolism	LC-MS		
61	Urea	AAs metabolism/Urea Cycle		NMR	5.884_5.732
62	Uridine	Nucleotide metabolism	LC-MS	NMR	7.893-7.863
63	Threonine	AAs metabolism	LC-MS		
64	Thymidine	Nucleotide metabolism	LC-MS		
65	Tryptophan	AAs metabolism	LC-MS		
66	Tyrosine	AAs metabolism	LC-MS		
67	Valine	AAs metabolism	LC-MS	NMR	1.043-1.036

[1] Fucose and methionine were not reliably detected in Chapter 4.

Abbreviations: AAs, amino acids; cAMP, 3',5'-cyclic adenosine monophosphate; BHB, β-hydroxybutyrate; CMP, cytidine monophosphate; FMN, flavin mononucleotide; Glu-1-P, glucose-1-phosphate; Gal-1-P, galactose-1-phosphate; HYP, hydroxyproline; Met (O), methionine sulfoxide; Nac-Gal, N-acetyl-galactosamine; Nac-Glu, N-acetyl-glucosamine; Nac-NA, N-acetyl-neuraminic acid; TMAO, trimethylamine N-oxide; UDP-Nac-Gal, uridine diphosphate-N-acetyl-galactosamine; UDP-Nac-Glu, uridine diphosphate-N-acetyl-glucosamine.

Table 4.1. The plasma metabolites detected either by Liquid chromatography-mass spectrometry (LC-MS), or by nuclear magnetic resonance (NMR), or by both.

No.	**Metabolites**	**Function**	**Detection by**		**Integration (ppm) in NMR spectrum**
1	Acetyl-carnitine	Lipid Metabolism	LC-MS		
2	Acetate	Energy Metabolism		NMR	1.938-1.912
3	Acetone	Lipid Metabolism		NMR	2.247-2.232
4	Alanine	AAs1 Metabolism	LC-MS	NMR	1.502-1.468
5	Allantoin	Energy Metabolism	LC-MS	NMR	5.422-5.370
6	Aminobutyrate	Energy Metabolism	LC-MS		
7	Arginine	AAs Metabolism/Urea Cycle	LC-MS		
8	Asparagine	AAs Metabolism	LC-MS		
9	Aspartate	AAs Metabolism/Urea Cycle	LC-MS		
10	BHB	Lipid Metabolism		NMR	2.333-2.288; 1.222-1.176
11	Carnitine	Lipid Metabolism	LC-MS		
12	Carnosine	AAs Metabolism	LC-MS		
13	Choline	Lipid Metabolism	LC-MS	NMR	3.212-3.203
14	Citrate	Energy Metabolism		NMR	2.701-2.649; 2.560-2.514
15	Citrulline	AAs Metabolism/Urea Cycle	LC-MS		
16	Creatine	Protein Degradation	LC-MS	NMR	3.944-3.932
17	Creatinine	Protein Degradation	LC-MS	NMR	4.069-4.058
18	Cystathionine	Nucleotide Metabolism	LC-MS		
19	Cytidine	Nucleotide Metabolism	LC-MS		
20	Epinephrine	Energy Metabolism	LC-MS		
21	Formate	Energy Metabolism		NMR	8.474-8.449
22	Glucose	Energy Metabolism		NMR	3.260-3.235
23	Glutamate	AAs Metabolism	LC-MS		
24	Glutamine	AAs Metabolism	LC-MS	NMR	2.481-2.444
25	Glycine	AAs Metabolism	LC-MS	NMR	3.571-3.561
26	Guanosine	Nucleotide Metabolism	LC-MS		
27	Hippurate	AAs/Energy Metabolism		NMR	7.861-7.822;7.667-7.625;7.585-7.537
28	Histidine	AAs Metabolism	LC-MS	NMR	7.800-7.775; 7.077-7.043
29	Hydroxyproline	AAs Metabolism	LC-MS		
30	Isoleucine	AAs Metabolism	LC-MS	NMR	1.026-1.005
31	3-methyl-KIV	AAs/Energy Metabolism		NMR	1.087-1.078
32	Kynurenine	AAs Metabolism	LC-MS		
33	Lactate	Energy Metabolism	LC-MS	NMR	4.133-4.112
34	Leucine	AAs Metabolism	LC-MS	NMR	0.970-0.950
35	Lysine	AAs Metabolism	LC-MS	NMR	3.037-3.018
36	Methanol	Energy metabolism		NMR	3.372-3.362
37	Methionine	AAs Metabolism	LC-MS		
38	Met (O)	AAs Metabolism	LC-MS		
39	Niacinamide	Energy Metabolism	LC-MS		
40	Norepinephrine	Energy metabolism	LC-MS		
41	Ornithine	AAs Metabolism/Urea Cycle		NMR	3.079-3.055

Table 4.1. (continued)

42	Pantothenate	Energy/Lipid Metabolism	LC-MS		
43	Phenylalanine	AAs Metabolism	LC-MS	NMR	7.457-7.414; 7.403-7.363; 7.351-7.319
44	Proline	AAs Metabolism	LC-MS		
45	Serine	AAs Metabolism	LC-MS		
46	Serotonin	AAs Metabolism	LC-MS		
47	Threonine	AAs Metabolism	LC-MS		
48	Thymidine	Nucleotide Metabolism	LC-MS		
49	Tryptophan	AAs Metabolism	LC-MS		
50	Tyrosine	AAs Metabolism	LC-MS	NMR	7.218-7.179; 6.929-6.881
51	Uric acid	AAs Metabolism	LC-MS		
52	Uridine	AAs Metabolism	LC-MS		
53	Valine	AAs Metabolism	LC-MS	NMR	1.066-1.078; 1.026-0.984
54	1,2-propanediol [1]	Lipid Metabolism		NMR	1.160-1.139
55	α-KIV [1]	AAs/Energy Metabolism		NMR	1.109-1.090
56	Isobutyrate [1]	Lipid Metabolism		NMR	1.066-1.058

[1] Isobutyrate 1,2-propanediol, α-ketoisovalerate were tentatively labeled in NMR spectra of blood.
Abbreviations: AAs, amino acids; BHB, β-Hydroxybutyrate; KIV, ketoisovalerate; Met (O), methionine sulfoxide.

Table 6.1. Loadings of first and second Principal Component (PC) of 5 plasma metabolite and metabolic hormone, which are obtained from Principal Component Analysis through lactation week 1 to 7 postpartum.

wk	**Loading score of 5 plasma metabolites and metabolic hormones** [1]									
	Free fatty acids		**BHB**		**Glucose**		**Insulin**		**IGF-1**	
	1st PC	2nd PC	1st PC	2nd PC	1st PC	2nd PC	1st PC	2nd PC	1st PC	2nd PC
1	-0.49	0.35	-0.36	-0.64	0.39	0.53	0.50	-0.14	0.48	-0.41
2	-0.47	0.16	-0.41	-0.75	0.47	0.23	0.46	-0.25	0.41	-0.55
3	-0.47	0.27	-0.40	-0.71	0.50	0.25	0.45	-0.10	0.40	-0.59
4	-0.49	0.29	-0.37	-0.75	0.49	0.29	0.46	-0.15	0.41	-0.5
5	-0.47	0.20	-0.39	-0.80	0.52	0.22	0.46	-0.27	0.42	-0.44
6	-0.47	0.20	-0.39	-0.81	0.52	0.17	0.46	-0.37	0.38	-0.38
7	-0.49	-0.21	-0.32	0.86	0.50	-0.07	0.49	0.46	0.41	-0.05

[1] Raw data is scaled by 4 times in Fig. 6.1 to make a visible bi-plot.

Table 6.2. The error rate (SD) of machine learning algorithms to predict metabolic status (good, average and poor) of dairy cows through lactation week 1 to 7.

Algorithms	**Error rate in each lactation week (%)**							*Average*
	week 1	week 2	week 3	week 4	week 5	week 6	week 7	
DT	38.5 (2.9)	24.3 (2.2)	28.1 (2.3)	24.9 (2.2)	26.0 (2.3)	28.2 (2.3)	25.0 (2.3)	27.9 (5.1)
NB	42.3 (1.8)	22.5 (1.2)	23.8 (1.7)	22.1 (1.5)	23.7 (1.6)	25.6 (1.3)	29.7 (2.0)	27.1 (6.8)
BN	42.5 (2.7)	22.0 (1.4)	27.4 (2.5)	23.1 (2.1)	25.7 (2.3)	27.9 (1.6)	25.3 (1.9)	27.7 (6.7)
SVM	33.6 (1.5)	22.5 (1.4)	21.3 (1.5)	17.3 (1.5)	22.5 (1.2)	24.3 (1.6)	23.0 (1.5)	23.5 (4.8)
ANN	48.7 (5.0)	31.9 (1.9)	32.4 (2.3)	33.9 (4.0)	35.5 (2.9)	35.7 (2.8)	42.7 (5.0)	35.3 (4.9)
Bootstrap	30.8 (1.9)	21.6 (1.4)	23.5 (1.5)	21.8 (1.5)	23.4 (1.4)	25.0 (1.5)	22.8 (1.8)	24.0 (3.2)
RF	30.8 (1.7)	22.1 (1.5)	22.6 (1.4)	20.2 (1.5)	23.8 (1.5)	24.7 (1.5)	24.3 (1.6)	24.1 (3.4)
KNN	36.7 (2.2)	29.6 (1.8)	28.3 (1.8)	25.3 (1.6)	32.3 (1.9)	30.6 (1.8)	34.5 (2.0)	31.0 (4.0)
Average	38.9 (5.0)	23.9 (3.4)	25.3 (3.5)	22.5 (3.8)	25.8 (4.0)	27.0 (3.1)	27.1 (5.2)	

Abbreviations: ANN, Artificial Neural Networks; BA, Bootstrap Aggregation; BN, Bayesian Network; DT, Decision Tree; KNN, K-Nearest Neighbour; NB, Naive Bayes; SVM, Support Vector Machine; RF, Random Forest.

Table 6.3. Plasma metabolite and metabolic hormone concentrations for dairy cows in 3 metabolic clusters in lactation week 1 to 7 postpartum. Cluster analysis of metabolic status was based on the concentration of plasma free fatty acids (FFA), BHB, glucose, insulin, and IGF-1. Values represent means (SD).

wk	Cluster	Dry period length			Parity	BW (kg)	MilkYield (kg)	MilkStd (kg)	Fat (%)
		0-d	30-d	60-d					
1	Good	33 (33.0%)	2 (3.1%)	-	2.9	652.9 (54.3)^{a}	14.1 (7.1)^{a}	4.13 (1.48)^{a}	5.20 (0.95)^{a}
	Average	53 (53.0%)	46 (71.9%)	36 (64.3%)	3.5	678.6 (74.0)^{b}	24.7 (6.7)^{b}	5.53 (2.13)^{b}	5.28 (1.04)^{a}
	Poor	14 (14.0%)	16 (25.0%)	20 (35.7%)	3.9	728.1 (59.4)^{c}	28.4 (6.7)^{b}	5.98 (2.88)^{b}	6.13 (1.32)^{b}
	P-value					0.01	< 0.01	<0.01	<0.01
2	Good	71 (56.3%)	8 (7.6%)	1 (1.2%)	2.8	653.8 (64.2)^{a}	25.1 (7.0)^{a}	1.59 (0.62)^{a}	5.03 (0.61)^{a}
	Average	48 (38.1%)	85 (81.0%)	66 (78.6%)	3.7	691.5 (71.7)^{b}	35.7 (5.6)^{b}	1.98 (0.78)^{b}	5.03 (0.74)^{a}
	Poor	7 (5.6%)	12 (11.4%)	17 (20.2%)	3.7	716.0 (63.5)^{b}	34.7 (6.4)^{b}	2.42 (0.97)^{c}	5.82 (0.93)^{b}
	P-value					< 0.01	< 0.01	<0.01	<0.01
3	Good	71 (59.2%)	19 (17.8%)	5 (6.0%)	2.8	669.7 (71.2)^{a}	30.5 (7.8)^{a}	1.19 (0.55)^{a}	4.70 (0.61)^{a}
	Average	43 (35.8%)	81 (75.7%)	71 (84.5%)	3.7	675.5 (65.2)^{a}	39.9 (6.2)^{b}	1.70 (0.92)^{a}	4.70 (0.64)^{a}
	Poor	6 (5.0%)	7 (6.5%)	8 (9.5%)	3.7	718.9 (55.2)^{b}	39.3 (5.2)^{b}	2.48 (1.87)^{b}	5.34 (0.74)^{b}
	P-value					< 0.01	< 0.01	<0.01	<0.01
4	Good	90 (74.4%)	30 (28.3%)	9 (10.8%)	3.0	675.5 (66.0)	33.0 (7.5)^{a}	1.33 (1.35)^{a}	4.62 (0.62)^{a}
	Average	29 (24.0%)	71 (67.0%)	68 (81.9%)	3.8	676.5 (65.7)	43.3 (6.3)^{b}	1.82 (1.19)^{a,b}	4.50 (0.65)^{a}
	Poor	2 (1.7%)	5 (4.7%)	6 (7.2%)	3.4	688.8 (43.1)	41.3 (5.5)^{b}	2.45 (1.97)^{b}	5.19 (1.05)^{b}
	P-value					0.78	<0.01	<0.01	<0.01
5	Good	81 (65.9%)	23 (23.7%)	6 (7.4%)	2.9	669.6 (63.0)	33.4 (7.7)^{a}	1.78 (1.43)^{a}	4.51 (0.55)^{a}
	Average	38 (30.9%)	65 (67.0%)	66 (81.5%)	3.7	674.5 (67.2)	43.9 (6.7)^{b}	1.23 (0.75)^{a}	4.31 (0.63)^{b}
	Poor	4 (3.3%)	9 (9.3%)	9 (11.1%)	3.9	695.3 (54.7)	40.8 (7.4)^{b}	2.43 (2.22)^{b}	5.08 (0.63)^{c}
	P-value					0.24	<0.01	<0.01	<0.01
6	Good	81 (65.3%)	27 (27.6%)	9 (11.7%)	2.9	672.5 (68.3)	33.7 (7.3)^{a}	1.20 (0.68)^{a}	4.48 (0.55)^{a}
	Average	40 (32.3%)	63 (64.3%)	60 (77.9%)	3.7	674.7 (59.8)	44.8 (6.1)^{b}	1.60 (1.07)^{b}	4.17 (0.56)^{b}
	Poor	3 (2.4%)	8 (8.2%)	8 (10.4%)	4.0	689.3 (47.6)	42.2 (6.9)^{b}	2.23 (1.19)^{c}	4.65 (0.42)^{a}
	P-value					0.56	<0.01	<0.01	<0.01
7	Good	88 (72.7%)	44 (44.4%)	12 (15.4%)	2.9	675.3 (65.9)	36.0 (7.9)^{a}	1.24 (0.69)^{a}	4.37 (0.59)^{a}
	Average	33 (27.3%)	50 (50.5%)	61 (78.2%)	3.9	677.6 (61.2)	44.3 (6.8)^{b}	1.75 (1.30)^{b}	4.15 (0.60)^{b}
	Poor	-	5 (5.1%)	5 (6.4%)	4.2	689.4 (40.8)	43.8 (6.4)^{b}	1.60 (0.52)^{b}	4.55 (0.23)^{a,b}
	P-value					0.58	<0.01	<0.01	0.02

Table 6.3. Plasma metabolite and metabolic hormone concentrations for dairy cows in 3 metabolic clusters in lactation week 1 to 7 postpartum. Cluster analysis of metabolic status was based on the concentration of plasma free fatty acids (FFA), BHB, glucose, insulin, and IGF-1. Values represent means (SD).

wk	Cluster	Lactose (%)	Protein (%)	Fat (kg)	Lactose (kg)	Protein (kg)	FPCM[1] (kg)	SCC (10^4/mL)
1	Good	3.99 (0.76)	5.72 (1.36)^{a}	0.74 (0.41)^{a}	0.59 (0.35)^{a}	0.74 (0.28)^{a}	18.4 (9.0)^{a}	1811 (2863)^{a}
	Average	4.20 (0.42)	4.63 (0.73)^{b}	1.29 (0.38)^{b}	1.05 (0.31)^{b}	1.12 (0.28)^{b}	31.7 (8.8)^{b}	814 (1686)^{b}
	Poor	4.13 (0.24)	4.37 (0.68)^{c}	1.70 (0.48)^{c}	1.17 (0.30)^{c}	1.23 (0.31)^{c}	38.7 (8.1)^{c}	231 (180)^{c}
	P-value	0.28	<0.01	<0.01	<0.01	< 0.01	<0.01	<0.01
2	Good	4.29 (0.39)^{a}	4.42 (0.29)^{a}	1.25 (0.35)^{a}	1.06 (0.26)^{a}	1.12 (0.34)^{a}	29.3 (7.8)^{a}	840 (1811)^{a}
	Average	3.79 (0.32)^{b}	4.46 (0.16)^{b}	1.79 (0.33)^{b}	1.35 (0.22)^{b}	1.60 (0.26)^{b}	40.8 (6.4)^{b}	202 (287)^{b}
	Poor	3.62 (0.21)^{b}	4.36 (0.21)^{c}	1.99 (0.39)^{c}	1.26 (0.26)^{b}	1.52 (0.32)^{b}	41.6 (8.9)^{b}	142 (186)^{b}
	P-value	0.01	< 0.01	< 0.01	<0.01	<0.01	<0.01	<0.01
3	Good	3.92 (0.38)^{a}	4.55 (0.20)^{a}	1.41 (0.32)^{a}	1.18 (0.24)^{a}	1.39 (0.37)^{a}	33.8 (7.5)^{a}	528 (1390)^{a}
	Average	3.46 (0.29)^{a}	4.57 (0.14)^{b}	1.87 (0.34)^{b}	1.38 (0.21)^{b}	1.83 (0.30)^{b}	43.5 (6.6)^{b}	219 (512)^{b}
	Poor	3.22 (0.17)^{b}	4.45 (0.18)^{c}	2.09 (0.39)^{c}	1.26 (0.17)^{a}	1.75 (0.25)^{a,b}	44.5 (6.7)^{b}	168 (290)^{a,b}
	P-value	0.01	<0.01	<0.01	<0.01	<0.01	<0.01	0.02
4	Good	3.75 (0.38)	4.58 (0.18)^{a}	1.50 (0.30)^{a}	1.22 (0.22)^{a}	1.51 (0.36)^{a}	35.7 (6.9)^{a}	275 (716)
	Average	3.26 (0.25)	4.59 (0.14)^{b}	1.93 (0.32)^{b}	1.41 (0.20)^{b}	1.99 (0.30)^{b}	45.3 (6.2)^{b}	193 (452)
	Poor	3.04 (0.24)	4.50 (0.16)^{c}	2.13 (0.47)^{b}	1.25 (0.15)^{b}	1.86 (0.26)^{b}	46.5 (5.9)^{b}	159 (226)
	P-value	0.17	<0.01	<0.01	<0.01	<0.01	<0.01	0.43
5	Good	3.72 (0.37)	4.59 (0.15)^{a}	1.49 (0.30)^{a}	1.23 (0.23)^{a}	1.54 (0.36)^{a}	36.1 (7.2)^{a}	251 (498)
	Average	3.23 (0.33)	4.60 (0.17)^{a}	1.88 (0.34)^{b}	1.41 (0.19)^{b}	2.03 (0.33)^{b}	45.1 (6.7)^{b}	242 (723)
	Poor	3.05 (0.29)	4.53 (0.14)^{b}	2.07 (0.43)^{c}	1.24 (0.22)^{a}	1.85 (0.35)^{b}	44.8 (8.3)^{b}	350 (687)
	P-value	0.12	<0.01	<0.01	<0.01	< 0.01	< 0.01	0.77
6	Good	3.66 (0.40)	4.61 (0.16)^{a}	1.49 (0.28)^{a}	1.22 (0.22)^{a}	1.55 (0.35)^{a}	35.9 (6.8)^{a}	257 (532)
	Average	3.21 (0.29)	4.61 (0.12)^{a}	1.86 (0.30)^{b}	1.43 (0.17)^{b}	2.07 (0.29)^{b}	45.3 (5.9)^{b}	173 (352)
	Poor	2.94 (0.18)	4.58 (0.15)^{b}	1.96 (0.36)^{b}	1.24 (0.20)^{a}	1.93 (0.31)^{b}	44.5 (7.3)^{b}	162 (202)
	P-value	0.68	<0.01	<0.01	<0.01	<0.01	<0.01	0.24
7	Good	3.61 (0.38)	4.62 (0.14)^{a}	1.55 (0.30)^{a}	1.28 (0.22)^{a}	1.66 (0.37)^{a}	37.6 (7.0)^{a}	170 (332)
	Average	3.15 (0.28)	4.6 (0.13)^{b}	1.82 (0.31)^{b}	1.39 (0.21)^{b}	2.04 (0.33)^{b}	44.2 (6.7)^{b}	233 (463)
	Poor	3.04 (0.20)	4.56 (0.1)^{b}	1.99 (0.30)^{b}	1.32 (0.15)^{b}	1.99 (0.28)^{a,b}	45.7 (5.5)^{b}	132 (87)
	P-value	0.32	<0.01	<0.01	<0.01	<0.01	<0.01	0.35

[1] FPCM: fat-and-protein-corrected milk.

^{a,b,c} Values within one lactation week of dairy cows in different metabolic status with different superscripts differ ($P < 0.05$).

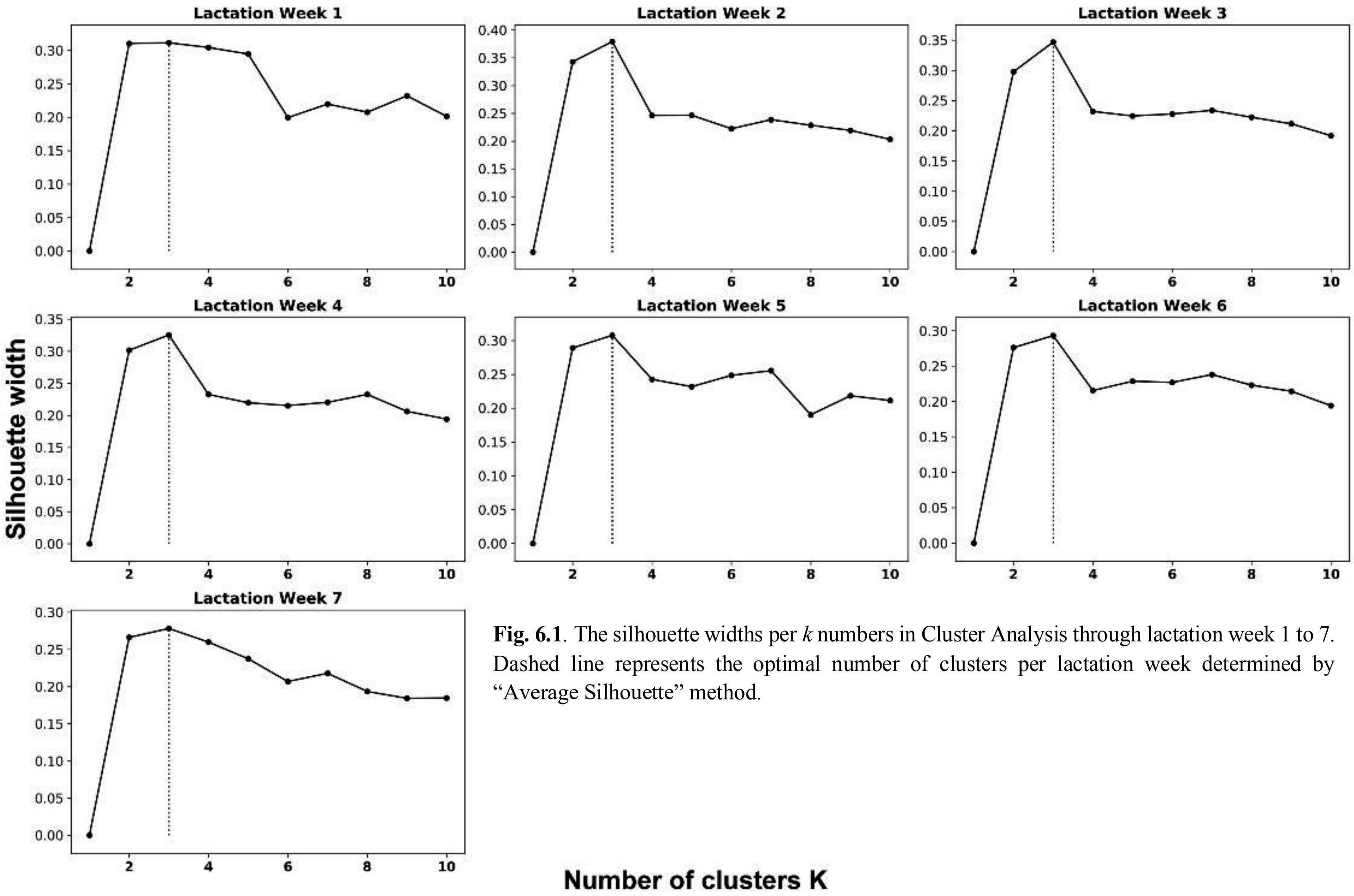


Fig. 6.1. The silhouette widths per *k* numbers in Cluster Analysis through lactation week 1 to 7. Dashed line represents the optimal number of clusters per lactation week determined by "Average Silhouette" method.

Summary

In early lactation, high-yielding dairy cows experience a negative energy balance (NEB), due to the increased energy demands for milk production and limited energy intake from feed. Negative energy balance is associated with impaired health and fertility. On farm, an early and reliable screening of energy balance and metabolic status should be able to identify dairy cows with an increased risk for health and fertility problems. Currently, energy balance and metabolic status can be estimated by milk characteristics, such as milk yield and milk fat, but also by the levels of ketone bodies in plasma or milk. The current screening methods are limited either by low estimated capability for estimating energy balance when using milk characteristics, or by the invasive sampling and labour-intensive requirements in specialized laboratories when using blood samples of dairy cows. Therefore, a non-invasive approach to estimate energy balance or metabolic status using milk samples or on-farm cow data could be interesting.

The aims of this thesis were, first to estimate energy balance and metabolic status of dairy cows using metabolomics and machine learning techniques; and second to investigate the metabolic pathways related to energy balance of dairy cows using metabolomics and proteomics techniques.

The first metabolomics study (Chapter 2) aimed to investigate metabolic profiles in milk through liquid chromatography mass spectrometry (LC-MS) and estimate energy balance using milk metabolites and milk production traits. In this study, 52 metabolites were detected from milk samples of 31 cows in lactation week 2 and 7 (Chapter 2). Partial least square (PLS) model could estimate energy balance as $Q^2 = 0.72$ and $Q^2 = 0.84$ for cows in lactation week 2 and 7, respectively. Through stepwise regression, reduced models could estimate energy balance of dairy cows in both lactation week 2 and 7, predictive capability ranged from 0.53 to 0.78, using milk production traits alone, or milk metabolites alone, or the combination of both. In these reduced models, milk fat yield, glycine, choline, and carnitine were important variables to estimate energy balance. The relationship of these milk metabolites with energy balance is proposed to be related to their roles in cell renewal.

The second metabolomics study (Chapter 3) aimed to measure the metabolic profile of milk of dairy cows in lactation week 2 through LC-MS and nuclear magnetic resonance spectroscopy (NMR), and to improve our understanding of metabolism in the mammary gland of dairy cows in early lactation. In this study, 67 metabolites were detected from milk samples of 87 cows in lactation week 2. Of 67 metabolites in milk, 25 were negatively correlated with energy balance, *r* ranged from -0.25 to -0.74; and 14 were positively correlated with energy balance, *r* ranged from 0.23 to 0.63. We concluded that the metabolic processes in the mammary gland during NEB are related to the leakage of cellular content, the synthesis of nucleic acids, the synthesis of cell membrane phospholipids, protein glycosylation, an increase in one-carbon metabolic processes as well as an increase in lipid-triglyceride anabolism.

The third metabolomics study (Chapter 4) aimed to *i*) reveal metabolic profiles of plasma and milk samples by integrating results of nuclear magnetic resonance (NMR) and liquid chromatography triple quadrupole mass spectrometry (LC-QQQ-MS); *ii*) investigate the relationship between energy balance and metabolic profiles of plasma and milk samples. In this study, 53 and 65 metabolites were detected from plasma and milk samples of 24 cows in lactation week 2, respectively. Total 10 metabolites presented a correlation between their concentration in plasma and in milk. Of the 53 metabolites in plasma 27 were correlated with energy balance. These plasma metabolites are related to mobilization from body fat, skeleton muscle, and bone, increased blood flow, and gluconeogenesis. Of the 65 metabolites in milk 30 were correlated with energy balance. These milk metabolites are related to the apoptosis and proliferation of cells in the mammary gland. In conclusion, metabolic profiles of plasma and milk clearly reflect the metabolism in the body or in the mammary gland of cows in NEB in early lactation.

The proteomics study (Chapter 5) is a preliminary study which aimed to investigate the correlation of energy balance with low abundant proteins in milk, as well as specific proteins with post translational methylation modifications (PTMMs). In this study, milk samples of 5 dairy cows in lactation week 2 were measured. The results indicate that energy balance is not only correlated with the 9 milk proteins concentration, but also correlated with 2 proteins with PTMMs. These milk proteins were related to immune response, lipid metabolism, and cell division in the mammary gland. Due to the limited sample size, the effect of limited sample size on statistical correlation could not be excluded. More milk proteins and specific proteins with PTMMs are expected to be correlated with energy balance in further study using a larger sample size than current study.

The first machine learning study (Chapter 6) aimed to *i*) investigate whether metabolic status of individual cows in early lactation could be clustered based on their plasma values, and *ii*) evaluate machine learning algorithms to predict metabolic status using on-farm cow data. First, dairy cows were clustered as either a good, or an average, or a poor metabolic status through 5 metabolic metabolites and metabolic hormones. Second, on-farm cow data, including dry period length, parity, milk production traits, and body weight, were used to predict metabolic status with 8 machine learning algorithms. Random Forest and Support Vector Machine were top 2 best performing algorithms to predict metabolic status. In Random Forest, milk yield, fat yield, protein percentage, and lactose yield had important roles in prediction, but their rank of importance differed across lactation weeks. In conclusion, dairy cows could be clustered for metabolic status, based on plasma metabolites and metabolic hormones. Moreover, on-farm cow data can predict cows in a good or an average metabolic status with best performance for Random Forest and SVM of all algorithms.

The second machine learning study (Chapter 7) aimed to *i*) evaluate if hyperketonemia in dairy cows (defined as plasma BHB ≥ 1.0 mmol/L) can be predicted using on-farm cow data either in current or previous lactation week, and *ii*) study if adding individual net energy intake (NEI) can improve the predictive ability of the model Partial least square discriminant analysis (PLS-DA) outperformed Random Forest, Artificial Neural Networks, and Support Vector Machine. When NEI was included in the model, the accuracy to predict hyperketonemia improved, but only to a limited extent. Besides NEI, body weight, milk fat and protein content and milk fat and protein yield were important variables to predict hyperketonemia, but their rank of importance differed across lactation weeks.

In the general discussion (Chapter 8), we first discussed the estimation of energy balance using metabolites in plasma and in milk obtained from metabolomics studies, and using machine learning algorithms based on on-farm cow data. Second, estimation of metabolic status and hyperketonemia using machine learning algorithms based on on-farm cow data is discussed. Third, the relation of energy balance and metabolic pathways of dairy cows, based on plasma and milk metabolomics studies in discussed. Fourth, future perspectives concerning metabolomics, proteomics, and machine learning techniques approached in dairy cows are discussed. Lastly, concluding remarks were presented (section 8.5).

In conclusion, the energy balance of dairy cows can be estimated by milk metabolites based on metabolomics study, and metabolic status can be estimated by machine learning algorithms using on-farm cow data. In early lactation, energy balance of dairy cows was related to milk and plasma metabolites associated with body reserve mobilization, apoptosis, cell proliferation, and synthesis of milk production.

Acknowledgements

My appreciation to you beyond any word.

First, I want to express my gratitude to China Scholarship Council that supported me to study aboard.

Second, I want to express my gratitude to my promoter (Bas Kemp) and supervisors (Ariette van Knegsel and Jacques Vervoort). You gave me chance to study in the Netherlands. Lady first. Ariette, you are my scientific mother. You taught me how to make scientific work effective, explainable, detailed, specific, and well-planned. Jacques, maybe I did not always understand your lessons, but you facilitated my lab-work and helped me a lot from work to life. Bas, you are an excellent example of being an approachable professor who is knowledgeable and well-known. I want to become a person like you. During the past four years, I really learnt a lot of things from all of you. Your lessons and wisdom will benefit me in all my life.

I would like to express my gratitude to our “WHY DRY” team, Ariette, Juncai, Novi, Renny, Akke, Eline and Junnan. Thanks for all the help from you, and I really miss our “WHY DRY coffee break”. Special thanks to Xin Zhu (Shenyang Agricultural University) and Juncai, you let me know Wageningen University and our group, ADP.

I would like to express my gratitude to all the collaborators for their valuable contribution to this thesis. Rupert Bruckmaier, Sjef Boeren, Edoardo Saccenti (Edo), and Renny van Hoeij, thank you for your continuing collaboration, supports, and comments on the chapters. Edo, thank you for your guidance and our discussions concerning machine learning algorithms and visualization techniques. Sjef, you taught me the basis of proteomics techniques, and I know I am still a beginner in the front of proteomics.

I also would like to express my gratitude to our secretarial team, Nanette van Hapert and Lora van der Kleijn. You paid a lot of time and patience to help me. Due to your help, I could pay all attention to this thesis.

I deeply appreciate Dr Ákos Kenéz (City University of Hong Kong), I am very happy to work with you!

Thanks all my Chinese friends. We didn’t know each other in China, but we had built friendship in Wageningen. Thank you all:

边梦溪、包曼及包伟、陈俊材及文婷婷、陈宝建、陈露、陈敏杰、陈琪婷、陈文祯、程旭、崔宏锐、邓昭举、冯涛及王晶、高保艳、高俊峰、何源、华登科、荀芳、古春凤、顾俊飞、顾生浩、关莜楠、姜倩倩、姜洋及段清瑶、蒋彬、李国华、李虎臣、李文峰、李鹤群、刘洁宇、刘琅青、刘志彬、路海博、罗露、罗琦、孟力、邵丽萍、邵子南及邵敏、石文标、万书明、王芳、王冠霖及汪奥嘉、王秋宇、王昭君、王宇蕴、王艳婷、王子申、于宏威、余甜、吴舟、夏淑雯、邢思远、熊玲、徐丹、徐芬、闫贞贞、杨宁、杨玉洁、袁桃林、张晨、张景艳、张仕仪、张伟、周礼斌、朱玉浩、朱聪、邹怡及吴佳。

另外，我特别感谢我的硕士导师单安山教授（东北农业大学）。恩师的教诲伴我左右，恩师对工作的孜孜不倦、在学术上的开拓创新鼓并激励着我。在此祝愿恩师身体健康，学术再攀高峰。

此处我想特别提及谈美秀女士。我深深地爱过你，也尝试过挽回，很遗憾没法继续曾经的爱情。但我必须承认，在完成此论文时候，你给予莫大的帮助。我只能由衷地祝福你有一个精彩的未来及幸福的生活。

最后，我要感谢我的父母及家人，给予了我生命、养育、及无限的支持。我自幼顽劣、不束管教，正是你们这三十年的陪伴及教诲让我一步步走到今天。从今往后，只愿你们能保养好身体，健康长寿，就是对我未来工作的最大支持了。

最后的最后，我只想以此怀念我遥在天堂的外婆，外婆是我生命中最疼我、最关爱我的人、是陪伴我时间最多的人。外婆在 2018 年 10 月永远地离开了我。外婆离世前的一年里，我感觉我一直在与癌细胞赛跑，不敢一丝懈怠，寝食难安。无论以后何如，外婆再也看不到今天、及未来的我了，每念于此，我的眼里就饱含泪水。逝者已去，愿外婆在天堂里没有病痛。我更要活好每一天，不会懈怠每一秒，陪伴好我的父母及外公。生命因此而美好，生命因此而伤感。

About the author

Curriculum Vitae

I (Wei Xu) was born 05 Oct. 1989 in Yichun, Heilongjiang, China.

I started BSc in Qingdao Agriculture University, China (09/2009 - 06/2012).

I started MSc in Northeast Agriculture University, China (09/2012 - 06/2015).

I started PhD in Wageningen University, the Netherlands (09/2015 - 08/2019).

Academic publications

***Xu W**, van Knegsel, A, Vervoort J, Bruckmaier R, van Hoeij, R., Kemp, B., & Saccenti E. (2019) Prediction of metabolic status of dairy cows in early lactation with on-farm cow data and machine learning algorithms. Journal of Dairy Science. 102(11): 10186-10201.*

***Xu W**, Vervoort J, Saccenti E, van Hoeij, R., Kemp, B., & van Knegsel, A. (2018) Milk Metabolomics Data Reveal the Energy Balance of Individual Dairy Cows in Early Lactation. Scientific Reports, 8(1): 15828.*

***Xu W.**, Zhu X., Tan T., Li W., Shan A. (2014) Design of Embedded-Hybrid Antimicrobial Peptides with Enhanced Cell Selectivity and Anti-Biofilm Activity. PloS One, 9(6): e98935.*

*Li, W., Tan, T., **Xu, W.**, Xu, L., Dong, N., Ma, D., & Shan, A. (2016). Rational design of mirror-like peptides with alanine regulation. Amino acids, 48(2), 403-417.*

*Zhu, X., Shan, A., Ma, Z., **Xu, W.**, Wang, J., Chou, S., & Cheng, B. (2015). Evaluation of the bactericidal efficiency and modes of action of a novel antimicrobial peptide T9W against Pseudomonas aeruginosa. Antimicrobial Agents and Chemotherapy, AAC-04830.*

*Zhu, X., Zhang, L., Wang, J., Ma, Z., **Xu, W.**, Li, J., & Shan, A. (2015). Characterization of antimicrobial activity and mechanisms of low amphipathic peptides with different α-helical propensity. Acta Biomaterialia, 18, 155-167.*

*Sun, J., Chen, Y., Qin, F., Guan, X., **Xu, W.**, & Xu, L. (2017). Prokaryotic expression of chicken interferon-γ fusion protein and its effect on expression of poultry heat shock protein 70 under heat stress. Animal Science Journal, 88(6), 882-892*

Training and supervision plan

Description	Year
Compulsory courses (3.0 ECTS)	
WIAS Introduction Course	2015
Course on philosophy of science and ethics	2018
Course on essential skills	2016
Writing of project proposal (6.0 ECTS)	
Non-invasive Biomarkers for Health and Metabolism in Dairy Cows	2016
Post-graduate courses (7.1 ECTS)	
Energy metabolism and body composition in nutrition and health research	2016
Specialist course on Lipids in Ruminants	2016
Metabolomics bioinformatics for life scientists	2017
Epigenesis & Epigenetics	2017
Proteomic course	2017
Structural Equation Modelling	2018
Competence and skills courses (8.0 ECTS)	
Advanced Statistics Course: Design of Experiments	2015
Statistics for the Life Sciences	2016
Scientific Writing	2016
Efficient Writing Strategies	2016
PhD Competence assessment	2016
Shaping future animal systems: exploring practices through dialogue	2017
Techniques for Scientific Writing and Presenting	2018
PhD discussion group (3.0 ECTS)	
WIAS Science Day Committee members	2017
WIAS science day (0.9 ECTS)	
WIAS Science day (2016, 2018, 2019)	-
Presentation Skills (4.0 ECTS)	
Poster presentation at ICPD conference, Wageningen, the Netherlands	2016
Oral presentation at 42th Animal Nutrition Research forum, Gent, Belgium	2017
Oral presentation at 6th BASF-Meeting, Viterbo, Italy.	2017
Oral presentation at WIAS Science Day, Wageningen, the Netherlands	2018

Colophon

This research described in this thesis was financially supported by Dutch Dairy Board (PZ, the Netherlands), Product Board Animal Feed (PDV, the Netherlands), and CRV (the Netherlands). Wei Xu was sponsored by a Chinese Scholarship Council (CSC) Fellowship.

Cover design:	Wei Xu and Yujie Yang
Thesis design and layout:	Wei Xu

Cover formatting and thesis printed by: ProefschriftMaken