Food Chemistry: X 5 (2020) 100080

Contents lists available at ScienceDirect

Food Chemistry: X

journal homepage: www.journals.elsevier.com/food-chemistry-x

An improved LC–MS method to profile molecular diversity and quantify the six main bovine milk proteins, including genetic and splicing variants as well as post-translationally modified isoforms

Guy Miranda[a,*], Leonardo Bianchi[a], Zuzana Krupova[a,1], Philippe Trossat[b], Patrice Martin[a,*]

[a] UMR GABI, INRAE, AgroParisTech, Université Paris-Saclay, 78350 Jouy-en-Josas, France
[b] ACTALIA, pôle expertise analytique, 39801 Poligny, France

ARTICLE INFO

Chemical compounds studied in this article:
Acetonitrile (PubChem CID: 6342)
Bis-Tris (PubChem CID: 81462)
Bronopol (PubChem CID: 2450)
DL-1,4-Dithiothreitol (PubChem CID: 446094)
Hydrochloric acid (HCl) (PubChem CID: 313)
Trifluoroacetic acid (TFA) (PubChem CID: 6422)
Tris buffer (PubChem CID: 6503)
Trisodium citrate dihydrate (PubChem CID: 71474)
Urea (PubChem CID: 1176)

Keywords:
Caseins
α-Lactalbumin
β-Lactoglobulin
Quantification
RP-HPLC
Mass spectrometry
Genetic polymorphisms
Phosphorylation

ABSTRACT

Here we describe a method based on Liquid Chromatography coupled with Mass Spectrometry (LC-MS) that provides an accurate determination of the six main bovine milk proteins, including allelic and splicing variants, as well as isoforms resulting from post-translational modifications, with an unprecedented level of resolution. Proteins are identified from observed molecular masses in comparison with theoretical masses of intact proteins indexed in an "in-house" database that includes nearly 3000 entries. Quantification was performed either from UV (214 nm) or mass signals. Thus, up to one hundred molecules, derived from the six major milk proteins, can be identified and quantified from an individual milk sample. This powerful and reliable method, initially developed as an anchoring method to estimate the composition of the six main bovine milk proteins from MIR spectra, is transferable to several mammalian species, including small ruminants, camels, equines, rabbits, etc., for which specific mass databases are available.

1. Introduction

Milk is not only a complete food, bringing essential nutrients such as proteins, fat, sugars, minerals, and micro-nutrients, it is also the vector of bioactive molecules that will ensure the neonate's development and growth whereas its immune system and defense capacities are, for most mammalian species, still immature at birth. There is indeed substantial evidence that milk contains many bioactive and health-promoting compounds affecting physiological functions or reducing disease risk. This statement is true for the main milk components, particularly for milk proteins (Korhonen & Pihlanto, 2007; Meisel, 2004). Numerous substantiated or potential bioactive protein components have been found, either as intact proteins or as derived peptides, encrypted in the protein sequences (Mohanty, Mohapatra, Misra, & Sahu, 2016). This is particularly true for the caseins that have long been claimed to be devoid of biological functions and designed only to ensure amino acid supply as well as phosphate and calcium absorption.

Milk is first consumed by neonates, however it is also widely consumed by all categories of consumers as such or after processing as milk-derived products, such as cheeses, fermented milk, etc. Here again, the caseins play a major role, since they are involved in the technological properties of milk, particularly in the milk-clotting process, a crucial step in cheese-making. The protein composition of individual milk samples with impaired coagulation or non-coagulation

* Corresponding authors.
E-mail addresses: guy.miranda@inrae.fr (G. Miranda), patrice.martin@inrae.fr (P. Martin).
[1] Present address: Excilone, Parc Euclide, 6 Rue Blaise Pascal, 78990 Elancourt, France.

https://doi.org/10.1016/j.fochx.2020.100080
Received 23 July 2019; Received in revised form 8 December 2019; Accepted 4 February 2020
Available online 17 February 2020

ability has been increasingly studied (Frederiksen et al., 2011; Ikonen, Ahlfors, Kempe, Ojala, & Ruottinen, 1999; Joudu, Henno, Värv, Kaart, & Kärt, 2007; Wedholm, Larsen, Lindmark-Månsson, Karlsson, & Andrén, 2006). However, the origin of this phenomenon is not yet fully understood, even though it seems that poorly- and non-coagulating milks contain a lower proportion of the 2 less-phosphorylated isoforms of αs-CN (αs1-CN 8P and αs2-CN 11P) and a lower proportion of glycosylated κ-CN (Jensen, Holland, Poulsen, & Larsen, 2012). The ability to quantify the casein content represents an issue of crucial importance for the dairy industry. Indeed, the natural variations in milk protein composition and concentration can markedly affect the yield of the cheese making process (Amalfitano et al., 2019; Wedholm et al., 2006), thus causing a direct and significant economic impact on the dairy industry.

Caseins, which represent more or less 80% of ruminant milk proteins, are essentially concentrated in the colloidal fraction of milk, in the form of highly hydrated and mineralized spherical particles: the so-called casein micelles. In cattle, caseins comprise a group of four peptide chains (αs1-, β-, αs2- and κ-CN) resulting from the expression of four tightly linked structural genes (*CSN1S1*, *CSN2*, *CSN1S2* and *CSN3*, respectively), of which the first three are evolutionary related (Rijnkels, 2002). Whereas the genomic organization of this locus is highly conserved across mammals, *CSN* genes have evolved rapidly to give rise to divergent proteins across mammalian species. The protein fraction of cow's milk is mainly composed (between 90 and 95%) of six major milk-specific proteins: four caseins (α_{s1}, α_{s2}, β and κ-CN) and two whey proteins: α-lactalbumin (α-LA) and β-lactoglobulin (β-LG), synthesized by the mammary epithelial cell.

The caseins (CN) are highly polymorphic in cattle and even more so in small ruminants. This feature is first due to the existence of numerous genetic variants (Martin, Bianchi, Cebo, & Miranda, 2013). In addition, the CN exhibit a high degree of heterogeneity due to post-translational modifications (PTM), mainly glycosylation (κ-CN) and phosphorylation (αs-, β- and κ-CN), which are critical for the formation and stability of CN micelles (Holland & Boland, 2014). Another feature, mainly regarding αs-CN, is the occurrence of splicing variants, arising from the usage of cryptic splice sites and from exon skipping (Martin, Cebo, & Miranda, 2013). These polymorphisms, and particularly those impacting the primary structure of the peptide chain, are not without consequences on the activity of peptides produced after digestion of caseins by proteases in the digestive tract of the consumer, whether adult or newborn. The presence of a wide range of bioactive peptides has been recently shown in the jejunal effluents of humans fed with milk proteins (Boutrou et al., 2013). It is therefore critical to be able to accurately establish the fine composition of the protein fraction of milks.

Methods conventionally used to profile milk proteins have long been based on gel electrophoresis techniques and liquid chromatography. Even though the Reverse Phase-High Performance Liquid Chromatography (RP-HPLC) proposed by Bobe, Beitz, Freeman, and Lindberg (1998) made it possible to separate and quantify (peak area from absorbance recorded at 214 nm) simultaneously the six major bovine milk proteins and some of their genetic variants, it did not provide any information on PTM, in particular on the phosphorylation level of CNs. Later Bordin, Cordeiro Raposo, De La Calle, and Rodriguez (2001) and more recently, Bonfatti, Grigoletto, Cecchinato, Gallo, and Carnier (2008), improved this method, using C4 or C5 columns instead of C18. Capillary Zone Electrophoresis (CZE) was also used for estimating the relative concentration of the individual main milk proteins (Heck et al., 2008). It was claimed that this method discriminates αs-CN differing in the phosphorylation states. However, some peaks are not resolved enough in CZE, making quantification difficult, and some genetic variants cannot be distinguished. This is the case for variants B and C of αs1-CN and variants E and A of κ-CN. Moreover, CZE separates κ-CN into multiple minor and one major peak (Miralles et al., 2001; Ortega, Albillos, & Busto, 2003), and minor peaks that represent different glycosylated or phosphorylated isoforms, co-migrate with β-CN A1 and A2 (Otte, Zakora, Kristiansen, & Qvist, 1997). In addition, identifying splicing variants that may represent in some species significant proportions of the cognate proteins is just impossible. Therefore, we have considered introducing mass spectrometry (MS) of which the ability to precisely identify and characterize proteins in complex mixtures has been demonstrated (Léonil et al., 1995; Mamone et al., 2003; Miralles, Leaver, Ramos, & Amigo, 2003), by coupling it with liquid chromatography (LC), thus providing a kind of second dimension of separation. However, although recent advances in the use of MS, in conjunction with protein/DNA sequence database search algorithms allow for identification, it has remained difficult to obtain accurate quantitative information despite recent efforts made to develop quantification methods based on the analysis of intact proteins using Extracted Ion Chromatograms (Vincent, Elkins, Condina, Ezernieks, & Rochfort, 2016).

The method that we describe here relies upon an LC-MS approach targeting native intact proteins. It provides a detailed and precise determination (qualitative and quantitative) of the six main bovine milk proteins including genetic and splicing variants, as well as isoforms resulting from PTM (phosphorylation, glycosylation). In addition, most of their main degradation products, in particular those resulting from proteolysis by plasmin, the endogenous milk protease, *i.e.* γ-caseins and their complements, can also be determined. In order to allow an automated identification of the six major milk proteins and their different isoforms as well as hydrolysis products, a library of theoretical masses was compiled for the bovine species. The method was validated by testing its linearity, reproducibility, repeatability, and accuracy. Since there is a growing interest for dairy species other than cattle, and even for model species such as humans and rodents, the method, first conceived and developed for bovine milk, is transferable to most mammalian species, including small ruminants, camels, equines, mice, and rabbits, for which specific milk protein mass databases have been created and implemented.

2. Materials and methods

2.1. Milk samples: collection and preparation

The milk samples were collected from cows of three different breeds: French Holstein-Friesian (FHF), Normande and Montbéliarde, either from INRAE experimental farms (Le Pin-au-Haras) or from herds of dairy producers in the Franche-Comté region (Montbéliarde breed). An optimal protocol of sample preparation was implemented to generate consistent and reliable data, minimizing the impact of various factors that influence data quality. Milk samples, preserved with bronopol and placed on ice immediately after milking, were rapidly skimmed by centrifugation at 2,500*g* and 4 °C for 20 min. The cream was removed by means of a spatula and the skimmed milk was aliquoted into fractions of 20 and 100 µL, frozen at −20 °C and kept under these conditions until analysis by Liquid Chromatography-Electro Spray Ionization-Mass Spectrometry (LC-ESI-MS).

2.2. Reverse-phase high performance liquid chromatography (RP-HPLC)

RP-HPLC was performed with an Ultimate LC 3000 system (Thermo Fisher Scientific, Waltham, MA) equipped with an auto sampler maintained at 15 °C and a dual wavelength detector (214 and 280 nm). The elution conditions were optimized to ensure the best separation of the six major bovine milk proteins and their main isoforms (genetic variants and PTM isoforms). Skim milk samples (20 µL) were clarified, at room temperature, by adding 180 µL of a clarification solution: 0.1 M Bis-Tris buffer (pH 8.0), 8 M urea, 4.4 mM trisodium citrate, and 19.5 mM DTT (Miranda, Mahé, Leroux, & Martin, 2004) to dissociate CN micelles. Clarified milk samples (20 µL) were injected directly into a Discovery BIO Wide Pore column C5 (150 × 2.10 mm, 300 Å), followed

by a wash of the injection needle with an (acetonitrile/water, 50/50, v/v) solution. The chromatographic conditions including the different steps of the elution gradient are given in Supplementary material (S0).

2.3. MS-parameters

The RP-HPLC output was directly interfaced with an ESI-TOF mass spectrometer micrOTOF II focus (Bruker Daltonics, Wissembourg, France). The positive ion mode was used and mass scans were acquired over a range of 50 to 3,000 *m/z*. End plate offset voltage was set at −500 V and capillary voltage to 4,500 V. Nebulizer gas (N_2) pressure was maintained at 250 kPa and drying gas (N_2) flow was set at 8.0 L/min at 200 °C. The LC-ESI-MS system was controlled by Hystar software v.2.3 (Bruker Daltonics). The charge number of multi-charged ions, the deconvoluted mass spectra, and the determination of average molecular masses (M_r) were obtained from Data Analysis v.3.4 software (Bruker Daltonics).

2.4. Identification of milk proteins

Identification of the major milk protein isoforms was achieved, thanks to the high mass accuracy of the TOF system, by comparing observed masses after deconvolution of the multi-charged ions spectrum, with theoretical masses of these proteins and their derivatives. In order to enable an automatic identification via the analysis software (Data analysis) we built a library of theoretical masses from literature data and genomic (NCBI) and protein (UniProtKB) sequence databases. In this database (Miranda, Bianchi, & Martin, manuscript in preparation) nearly 3000 theoretical masses corresponding to the multiple isoforms: genetic and splicing variants, PTM status and main known plasmin hydrolysis products of the six major bovine milk-specific proteins (*i.e.* αs1-, αs2-, β- and κ-CN, α-LA and β-LG)) are indexed.

2.5. Relative quantification of bovine milk proteins and isoforms

2.5.1. Quantification from absorbance at 214 nm

This approach is based on the integration of peak area (absorbance at 214 nm) to evaluate the proportions of each of the six milk-specific protein families, relative to the total integrated area of each chromatographic profile of individual milk samples. Corrective factors were introduced for each protein family to take into account specific absorbance at 280 nm due to their content in aromatic amino acids. In addition, milk proteolysis and particularly the proteolysis of αs1- and β-CN can be taken into account by quantifying specific degradation products, making it possible to estimate more precisely the relative proportions of each casein, after their reassignment. αs2-CN and β-CN have been reported to carry over from one injection to the following one on C18 columns (Nieuwenhuijse, van Boekel, & Walstra, 1991; Visser, Slangen, & Rollema, 1991). Although the column used here was a C5 (less hydrophobic than C18), we searched for a possible memory effect of the column. A milk sample consisting of the 50/50 blend of 2 individual FHF milks (cows #55 and #56) was injected onto the column using the optimized elution gradient followed by a regeneration step (10 min to 95% acetonitrile) and an equilibrating step under the initial conditions (10 min to 29.5% acetonitrile). Two successive blanks (clarification solution) were then performed under the elution conditions used to analyze the mix of the two milks. This sequence was reproduced three times.

2.5.2. Quantification using the intensity of the mass signal (IMS)

This method is based on the use of the IMS after deconvolution of the multi-charged ions spectrum, taking into account the ability of the different molecules to ionize. Therefore, the various factors expected to impact the ability to ionize (PTM (glycosylation, phosphorylation) and genetic variants) were evaluated.

To evaluate the effect of phosphorylation levels of proteins on ionization ability, we compared the IMS of a native casein (phosphorylated) to that obtained with the same totally dephosphorylated protein after treatment with the Calf Intestinal alkaline Phosphatase (CIP, SIGMA P6774; 50 units/μL). The conditions used to dephosphorylate CN are described in Supplementary material (S0).

Regarding the possible effect of genetic polymorphisms on ionization ability, the ratio Absorbance at 214 nm/Intensity of the deconvoluted Mass Signal (UV/IMS) was determined for the most frequent genetic variants of the same protein. The FHF milk samples analyzed (n = 39, in duplicate) came from animals chosen for their representativeness (most frequent genetic variants in the populations analyzed): *i.e.* β-CN A1, A2, A3, I and B, αs1-CN B and C, κ-CN A, E and B and variants A and B for β-LG. Since αs2-CN and α-LA are essentially monomorphic, they were not considered.

2.6. Qualification of the LC–MS method

Qualification (performance characteristics) of the LC-MS method was carried out, based on the current normative standards in this field, using FHF milks. The following criteria were studied: quantitative linearity (2 individual milks), intra-laboratory precision (repeatability and reproducibility, with 43 and 10 individual milks, respectively) and accuracy (1 milk spiked with 3 purified proteins).

2.6.1. Quantitative linearity

Quantitative linearity was estimated for the six major bovine milk proteins, using a range of milks with 11 levels of protein concentration, produced by mixing an ultrafiltration retentate (87.54 g/L) and a permeate applying a volume/volume dilution principle, by corrected density weighing (to obtain linear dilution coefficients). Each sample was analyzed by the LC-MS method in duplicate. For each sample, the theoretical total protein content was calculated, and a theoretical milk protein level was calculated for each protein (κ-CN, αs1-CN, αs2-CN, β-CN, α-LA and β-LG) by applying an average composition factor observed in a previous study (respective average % of the different milk proteins).

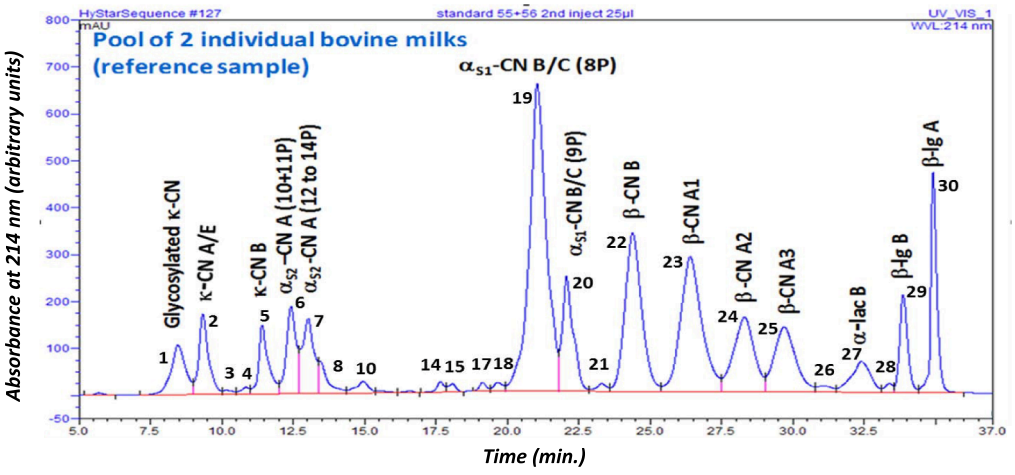
2.6.2. Intra-laboratory precision

Intra-laboratory precision was determined by evaluating the repeatability on 43 individual FHF cow's milk samples (duplicate analyses of two independent preparations for each sample) and intra-laboratory reproducibility through the analysis of 10 duplicate samples over 5 days (for a total of 200 measurements). The following parameters were calculated both in relative percent of total protein (%) for each milk protein and in g milk protein per liter of milk (g/L): Sr (standard deviation of repeatability) and SR (standard deviation of reproducibility) within the laboratory and r (maximum deviation between replicates) and R (maximum reproducibility deviation).

2.6.3. Determination of the accuracy

The accuracy of the method was determined by supplementation (spiking), at six different levels (0 to 75 μL, by increment of 15 μL), of a reference milk with preparations of three proteins (αs1-CN B-8P and 9P, β-CN A2-5P and α-LA) purified in our laboratory. Solutions of purified proteins for spiking were prepared by dissolving lyophilized proteins in water at the following concentrations: 0.725, 0.118 and 0.157 μg/μL, respectively, and their purity estimated in RP-HPLC at 90, 95 and 95%, respectively. Five replicates of six spiked milks, corresponding to six levels of supplementation, were analyzed, *i.e.* 30 measurements.

To convert relative into absolute quantification we used the Amido Black method, which is a colorimetric chemical method to quantify true protein content in milk. This method, standardized by AFNOR under the identification NF V 04 216, gives equivalent results to the Kjeldahl method (TN – NPN) × 6.38.



B

Peak number	Retention Time	Main masses Observed	Identification	Therotical masses	Area	Relative Area	Others	Protein family	Relative concentration
	min	Da		Da	mAU*min	%			%
1	8.46	20901.6615 20931.8947 21850.3951	κ-CN B 1P 2OG κ-CN A 1P 2OG κ-CN E 1P 3OG	20901.098 20933.054 21850.869	60.187	2.97		Glycosylated κ-CN	**2,97**
2	9.34	19007.3305 19037.2153	κ-CN E 1P κ-CN A 1P	19007.345 19037.371	66.752	3.30		κ-CN	**3.30**
3	10.15		n.a.		3.133	0.15	0.15		
4	10.84		n.a.		4.843	0.24	0.24		
5	11.41	19005.3940	κ-CN B 1P	19005.416	56.887	2.81		κ-CN	**2.81**
6	12.42	25148.4917 25228.5579	αs2-CN A 10 & 11P	25148.348 25228.328	74.018	3.66		αs2-CN	**8,43**
7	13.02	25308.4656	αs2-CN A 12P	25308.308	77.015	3.81			
8	13.43	25388.2338 25468.6767	αs2-CN A 13 & 14P	25388.288 25468.268	19.390	0.96			
9	13.89		n.a.		5.193	0.26	0.26		
10	14.95	12096.3734 12177.1687	PP5 f(1-105) A2/A3 4P PP5 A2/A3 5P	12097.294 12177.274	14.213	0.70		β-CN	**0.70**
11	15.58		n.a.		3.396	0,.17	0.17		
12	16.60		n.a.		3.0344	0.15	0.15		
13	17.38		n.a.		1.450	0.07	0.07		
14	17.66	21997.0835 22068.8598	αs1-CN C 8P Δe4 αs1-CN B 8P Δe4	21996.831 22068.910	7.395	0.37		αs1-CN	**0.75**
15	18.07	14031.9655 14105.6761	f (80-199) αs1-CN C 1P f (80-199) αs1-CN B 1P	14035.772 14107.836	7.746	0.38			
16	18.64		n.a.		1.598	0.08	0.08		
17	19.14	22426.9941 22601.9471	αs1-CN B 8P Δe8 ΔQ59 & Q78 αs1-CN B 7P Δe8	22427.422 22603.703	8.347	0.41		αs1-CN	**0.41**
18	19.66		n.a.		9.622	0.48	0.48		
19	21.04	23542.3689 23614.8144	αs1-CN C 8P αs1-CN B 8P	23542.648 23614.712	462.326	22.84		αs1-CN	**28.48**
20	22.07	23624.2794 23695.6948	αs1-CN C 9P αs1-CN B 9P	23622.628 23694.692	114.171	5.64			
21	23.29		n.a.		9.975	0.49	0.49		
22	24.39	24092.5726	β-CN B 5P	24092.319	237.283	11.72		β-CN	**37.91**
23	26.40	24023.4698	β-CN A1 5P	24023.209	257.687	12.73			
24	28.31	23983.5160	β-CN A2 5P	23983.185	138.520	6.84			
25	29.69	23974.1909	β-CN A3 5P	23974.174	124.969	6.17			
26	31.09	11558.4804	γ3 f (108-209) β-CN 0P	11558.612	9.201	0.45			
27	32.40	14185.8586	α-lactalbumin B	14186.064	53.580	2.64		α-LA	**2,65**
28	33.37	18605.2732	lactosyl-β-lactoglobulin B	18605.320	6.831	0.34		β-LG	**9,49**
29	33.86	18281.1176	β-lactoglobulin B	18281.208	68.804	3.40			
30	34.91	18367.2233 18691.6191	β-lactoglobulin A lactosyl-β-lactoglobulin A	18367.298 18691.300	116.472	5.75			
Total					2024.040	**100.00**	**2.10**		**97,90**

Fig. 1. Profiling of the six major milk protein families of cow's milk: identification by mass spectrometry and quantification from absorbance at 214 nm. Separation (A), identification and UV quantification (B) of bovine milk proteins including genetic variants and phosphorylation isoforms by RP-HPLC coupled with a microTOF Mass Spectrometer (Bruker Daltonics) from a pool of two reference milk samples (cows #55 and #56) of known genotypes: *AA-AA-BC-A2A3-BB-AB* and *EB-AA-BB-BA1-BB-AB*, respectively at the *CSN3*, *CSN1S2*, *CSN1S1*, *CSN2*, *LALBA* and *PAEP/BLG* loci. Identification (B) was confirmed from observed masses by comparison to theoretical masses of the known genetic variants. Relative quantification (B) was determined as relative area (%). Relative areas corresponding to non-identified masses (n.a.) are gathered as "others". Proteins of the same family were grouped together by adopting the following specific colour code: κ-CN in orange, darker orange for glycosylated κ-CN (peak 1); αs2-CN in grey; β-CN and derived molecules γ3 and PP5 in blue; αs1-CN in green, α-LA in pink and β-LG in yellow. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

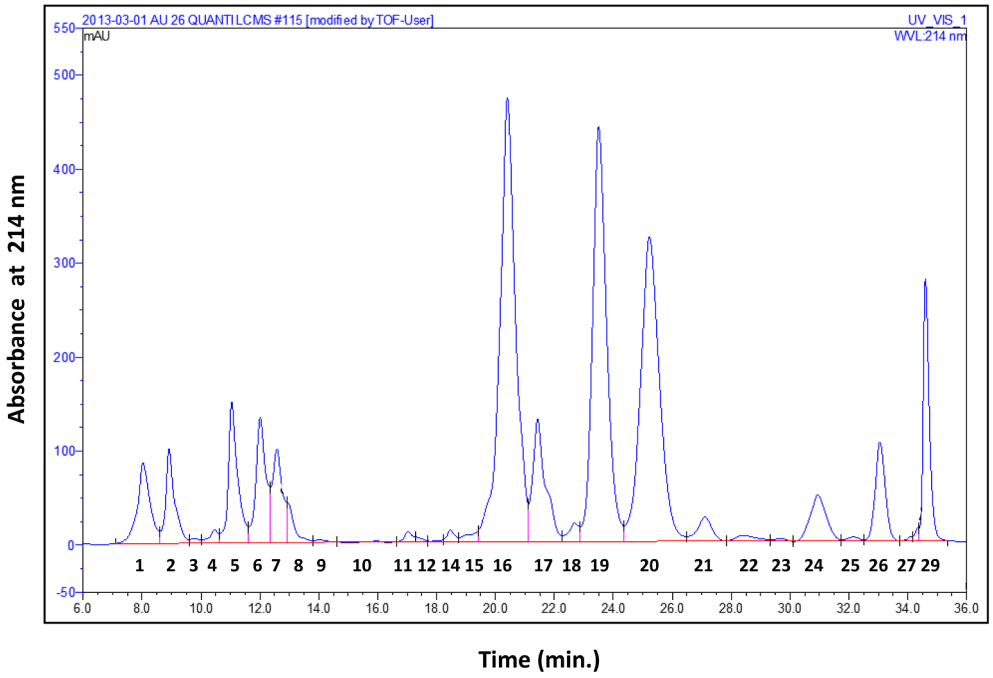


Fig. 2. RP-HPLC profile of an individual bovine milk sample. The chromatogram profile was partitioned into 29 peaks whose contents were identified from the data generated by the mass spectrometer (Table 1).

3. Results and discussion

3.1. Qualitative determination of milk proteins

3.1.1. Separation and identification of the main milk proteins and related genetic variants

The RP-HPLC that we developed and we report here is based on previously published works (Bobe et al., 1998; Jaubert & Martin, 1992; Miranda et al., 2004; Visser et al., 1991). A preliminary version of the method was first discussed (Miranda, Krupova, Bianchi, & Martin, 2013) and since then optimized to reach a high resolution level, allowing the discrimination of most of the genetic variants, PTM isoforms, as well as splicing variants and several proteolysis products, including γ-CN and their complements.

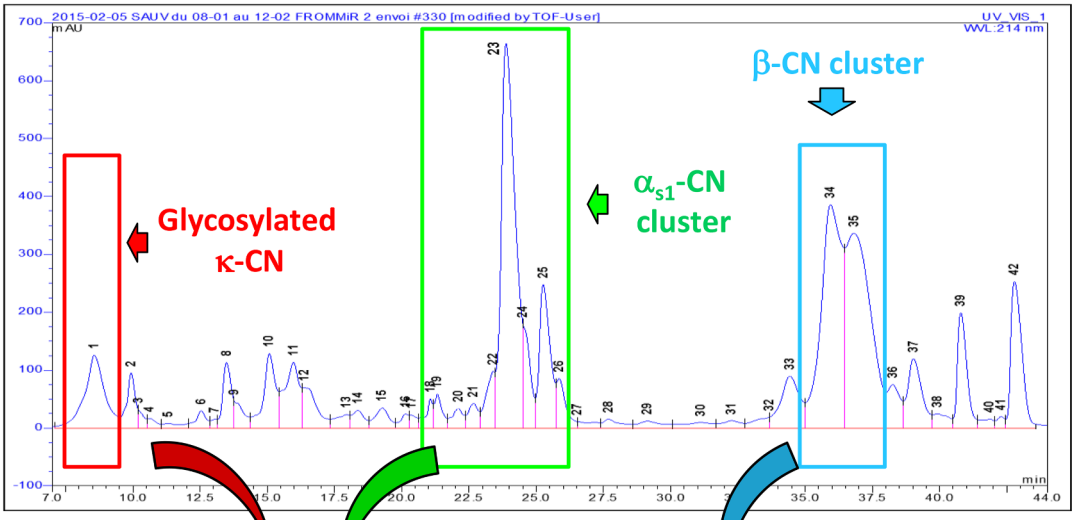
The identification, based on a comparison of the masses observed with theoretical molecular masses, deduced from amino acid sequence of proteins and known PTM, was validated by the analysis of milk standards from cows of known genotypes. Fig. 1A shows the elution profile of the major bovine milk proteins yielded from the pool of two individual skimmed milk samples used as standards: cow#55 and cow#56 whose genotypes were AA-AA-BC-A2A3-BB-AB and EB-AA-BB-BA1-BB-AB, respectively at the *CSN3*, *CSN1S2*, *CSN1S1*, *CSN2*, *LALBA* and *PAEP/BLG* loci. Such a sample shows the resolving power of the Biodiscovery C5 column, in the optimized chromatographic conditions used. The chromatographic profile was split into 30 peaks. A perfect resolution of four out of the five most frequent β-CN variants (A1, A2, A3, B and I) found in the populations studied, was thus achieved. β-CN variant I which is rather frequent in several populations, including the Italian (Jann et al., 2004) and Dutch (Visker et al., 2011) HF, Italian Simmental (Bonfatti et al., 2008), as well as the French Montbéliarde (Fang et al., 2016), co-eluted with variant A2. Its frequency can reach up to 20% in Dutch HF (Visker et al., 2011). To distinguish variant I from variant A2 it is therefore necessary to use mass data observed under the relevant compound. The difference in mass between the two variants is such (23,983.18 Da vs. 23,965.15 Da, for β-CN A2-5P and β-CN I-5P, respectively) that, after deconvolution of the multi-charged ions spectrum corresponding to the relevant peak, it becomes easy to determine which variant(s) we are dealing with (Fig. 3).

As far as κ-CN is concerned, it is necessary to underline that the major glycosylated isoforms of κ-CN are the very first proteins eluted as a single peak, regardless the variant. By contrast, non-glycosylated A and B variants are easily distinguished (Fig. 1A). However, it is again necessary to use mass information (Fig. 1B) to discriminate between non-glycosylated κ-CN A (19,037.37 Da) and E (19,007.34 Da) which are co-eluted in peak 2.

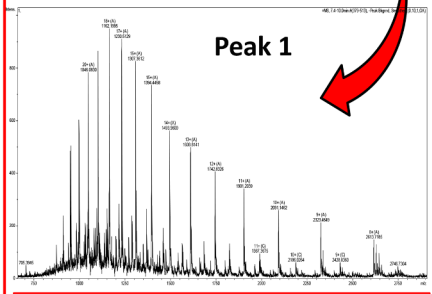
Likewise, the identification of αs1-CN variants B and C, which are co-eluted (Fig. 1A, peak 19) arises from the masses observed: 23,614.8144 and 23,542.3689 Daltons, respectively (Fig. 1B) compared with the theoretical values of 23,614.712 and 23,542.648, for their major isoform, with eight phosphate residues. It is worth noting that this is virtually the only way to distinguish, in such conditions, the αs1-CN B and C variants that differ only by the E192G mutation. Indeed, to our knowledge, there is no RP-HPLC method so far published resolving these two αs1-CN genetic variants (Bobe et al., 1998; Groen, van der Vegt, van Boekel, de Rouw, & Vos, 1994; Nieuwenhuijse et al., 1991; Visser et al., 1991). By contrast, Recio, Pérez-Rodríguez, Ramos, and Amigo (1997) claim that αs1-CN variants B and C are easily identified, at the heterozygous phenotype in CZE. The accuracy (difference between the observed mass and the theoretical mass) was between 2 and 15 ppm (± 0.02 and 0.35 Dalton on average) for the main isoforms (mass signal intensity > 5000).

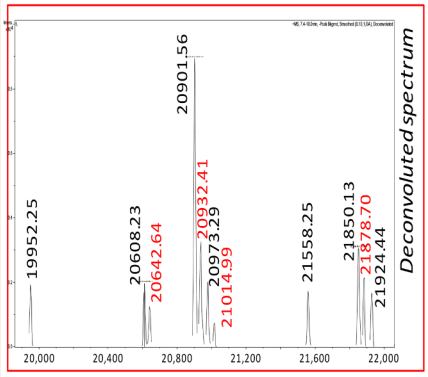
The most two frequent genetic variants (A and B) of β-LG are eluted at the end of the chromatogram (Fig. 1A) and identified, in peaks 30 and 29, based on their molecular masses: 18,367.30 and 18,281.21 Da, respectively (Fig. 1B). A lactosyl-β-LG conjugate (M_r = 18,605.27 Da), due to a covalently bound lactose residue, was also detected in peak 28 for the B variant, whereas the lactosyl conjugate of the A variant was co-eluted with the native β-lactoglobulin A, in peak 30. Free NH2 groups of basic amino acids, identified by Léonil et al. (1995) as K47 in bovine lactosyl-β-LG, react with the reducing carbonyl group of lactose forming the so-called Amadori products.

Although αs2-CN (peaks 6–8) and α-LA (peak 27) show genetic polymorphisms (Martin, Bianchi et al., 2013), these two proteins were quite monomorphic: variants A and B, respectively, in the bovine populations studied. However, αs2-CN displays a high molecular diversity due to its phosphate content.



Spectrum of multi-charged ions

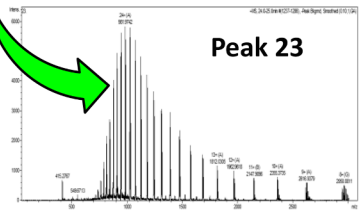




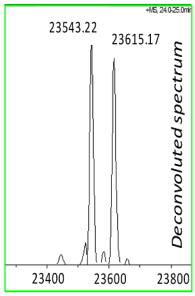
Glycosylated κ-CN A & B

observed masses	intensities	identification	glycosylation		
			GaN	Ga	SA
19952.25	1934	varB 1P 1xOG	1	1	2
20608.23	1990	varB 1P 2xOG	2	2	3
20642.64	1401	varA 1P 2xOG	2	2	3
20901.56	8952	varB 1P 2xOG	2	2	4
20932.41	3697	varA 1P 2xOG	2	2	4
20973.29	2070	varB 1P 3xOG	3	3	3
21014.99	736	varA 2P 2xOG	2	2	4
21558.25	1735	varB 1P 3xOG	3	3	5
21850.13	3070	varB 1P 3xOG	3	3	6
21878.70	2161	varA 2P 3xOG	3	3	5
21924.44	1761	varB 1P 4xOG	4	4	5

Spectrum of multi-charged ions

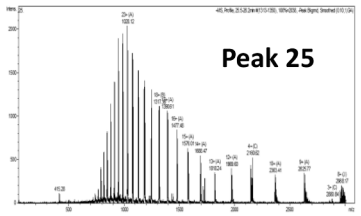


Masses	Intensities
22604.65	8166
22634.13	9635
22737.18	5631
23445.35	3075
23522.43	6571
23543.22	68187
23615.17	60222

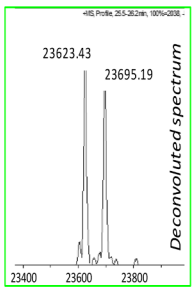


αs1-CN B & C - 8P

Spectrum of multi-charged ions

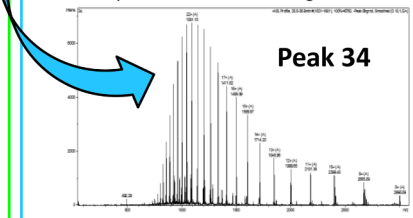


Masses	Intensities
22618.02	2290
23603.16	2561
23623.43	21908
23656.50	833
23675.88	1433
23695.19	20510
23717.09	1085

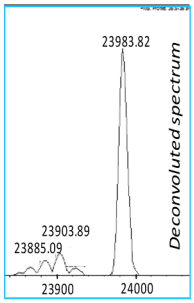


αs1-CN B & C - 9P

Spectrum of multi-charged ions

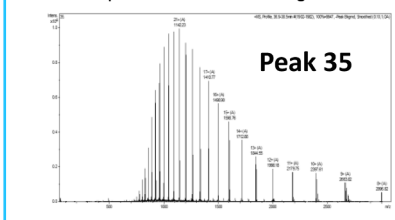


Masses	Intensities
20505.30	10943
20523.05	8598
23866.17	3160
23885.09	5863
23903.89	8213
23983.82	82796
24904.90	5688



β-CN A2 - 5P

Spectrum of multi-charged ions



Masses	Intensities
17669.79	847
22704.33	8848
23926.91	1158
23947.63	921
23965.85	117326
24080.16	2041

β-CN I - 5P

(caption on next page)

Fig. 3. Improved resolution of bovine milk proteins by RP-HPLC using an extended acetonitrile gradient. Milk was sampled from a cow of the following genotype: *AB*, *AA*, *BC*, *A2I*, *BB* and *AB,* respectively, at the *CSN3, CSN1S2, CSN1S1, CSN2, LALBA, BLG/PAEP* loci. The chromatogram profile (upper panel) was partitioned into 42 peaks whose contents were identified from the data generated by the mass spectrometer (Supplementary material S5). The glycosylated κ-CN (peak 1, framed in red), the αs1-CN cluster (framed in green), including peaks 22 to 26 and the β-CN cluster (framed in blue), including peaks 33 to 36, were subjected to an in-depth MS analysis to identify and quantify from deconvoluted mass signals, genetic variants and isoforms with different levels of PTM. κ-CN variants A and B and their glycosylated isoforms present in peak 1 were distinguished and quantified and the composition in *N*-acetyl galactosamine (GaN), Galactose (Ga) and Sialic Acid (SA) of their glycosylation units deduced. αs1-CN variants B and C, with 8P (peak 23, B) and 9P (peak 25, C) were easily distinguished and quantified. Masses and intensities corresponding to different isoforms are indicated with green arrows. β-CN variants A2-5P (peak 34, D) and I-5P (peak 35, E) were as well easily distinguished and quantified. Masses and intensities corresponding to β-CN allelic variants are indicated with blue arrows. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.1.2. Discrimination of isoforms arising from post-translational modifications and splicing anomalies

As shown in Fig. 2 and Table 1 (analysis of an individual FHF milk sample), a first resolution level of αs2-CN isoforms was obtained according to its number of phosphate groups thanks to the high efficiency of the chromatographic separation. Indeed, peaks 6, 7 and 8 correspond to isoforms 10-11P, 12P, and 13-14P, respectively. However, using mass data, the resolution is refined to distinguish a minimum of seven isoforms, ranging between 8 and 14 phosphate groups (Supplementary material S1). In some milk samples, an isoform with up to 15P was found (Fang et al., 2016).

With regards to glycosylation, which here mainly concerns κ-CN, we have seen that most of the glycosylated κ-CN isoforms were eluted as a single peak (peak 1) at the beginning of the chromatogram (Fig. 1A). However, some glycosylated κ-CN isoforms can be found all along the chromatogram (see Table 1 and § quantification). The detailed analysis of the multi-charged ions spectrum provides, after deconvolution, a series of masses corresponding to different O-glycosylation levels (see Fig. 3).

We also observed the existence of minor non-allelic isoforms of αs1-CN (Table 1, within peaks 11 and 14), with masses matching that displayed by genetic variants A (skipping of exon 4) and H (skipping of exon 8). In addition, isoforms corresponding to αs1-CN having lost Q59 and/or Q78 residues (Table 1, peaks 14 and 17) due to the use of cryptic intra exon splice sites (CAG codon at the 5′ end of the exon) during the course of messenger RNA processing, were detected. This phenomenon is well known and documented, particularly with regard to αs1-casein. It has been reported in almost all mammalian species including cattle and water buffalo (Martin et al., 2013). An additional isoform, hitherto unknown, displaying a putative deletion of exon 12 was also found in peak 17. Therefore, even if this kind of phenomenon is especially frequent with other species, including small ruminants, non-allelic splicing variants also occur in cattle and concerns, to a lesser extent, all three Ca-sensitive CN. Indeed, a mass corresponding to the αs2-CN D variant, characterized by a defective splicing of the precursor to mRNA leading to the loss of exon 8 (Martin, Bianchi et al., 2013), was detected among masses present in peak 5, in some individuals. Similarly, a β-CN non-allelic variant having lost the peptide sequence (ESITRINK), encoded by exon 5, was detected in peak 19 (Table 1). These exon-skipping phenomena were confirmed for both αs1-CN and β-CN in LC–MS/MS (Supplementary material S2).

By enriching the theoretical mass database with masses of hydrolysis products from caseins by plasmin, it becomes possible to extend the analysis to the identification and relative quantification of molecules such as γ-CN and their complements (*e.g.* PP5), as well as peptides derived from αs1-CN. Thus, in peak 10 (Fig. 1B) we identified peptides corresponding to the proteose-peptone component 5 (PP5, f(1–105) β-CN A2/A3) with 4 and 5 phosphate residues (with observed masses of 12,096.3734 and 12,177.1687 Da). As a consequence, γ3-CN, f (108–209) β-CN, was detected in peak 26 as a molecule of M_r 11,568.48 Da.

3.2. Quantitative determination of milk proteins

3.2.1. Recording absorbance at 214 nm to target quantification of milk protein families

Proteins identified under each of the major chromatographic peaks, defined by recording the absorbance at 214 nm, belong to the same protein family and account for most of their surface area. It is therefore possible to make a relative quantification of each peak of the chromatogram by expressing its surface as a percentage of the total area of the peaks of the chromatogram. The absorbance of a protein at 214 nm is generally considered as being proportional to the number of peptide bonds it contains. However, its content in aromatic amino acid residues may impact the measurement. Thus, to overcome such a pitfall, corrective factors were introduced, taking into account the content in aromatic amino acids of each protein, based on its molar extinction coefficient (ε) at 280 nm. Since the number of peptide bonds per gram of protein is essentially the same for all the main milk proteins, the area of each peak of the chromatogram, expressed as a percentage of the total area of all peaks, allows a relative quantification of each protein. Thus, the amount of each major bovine milk protein resolved in RP-HPLC under the conditions used (Fig. 1A), including genetic variants and differently phosphorylated isoforms, expressed in percent, is given in Fig. 1B as relative area. Because of their distribution in more than one peak, to get an individual protein composition, *i.e.* to quantify the four CN and the main two whey proteins (α-LA and β-LG) of the milk sample, it is possible to integrate peaks relevant to the same protein family. In such a way, the relative amount of αs2-CN, regardless of its phosphorylation levels, was estimated (8.43%) by summing the areas of peaks 6, 7 and 8. Similarly, peaks 14, 15, 17, 19 and 20, which correspond to different αs1-CN B and C isoforms, as well as peaks 22–26 which correspond to the four β-CN genetic variants, can be combined to assess the overall αs1-CN (29.64%) and β-CN (38.61%, including γ3-CN and its complements, in particular PP5) relative contents, respectively, in the milk sample. With regards to κ-CN, it is possible to aggregate or distinguish glycosylated and non-glycosylated κ-CN, although glycosylated κ-CN seems to be underestimated (see thereafter).

Therefore, the average distribution of the different protein families could be estimated. They were determined on a panel of 240 individual milk samples from Montbéliarde cows, compared with previous data available in the literature (Table 2). Interestingly, the introduction of the above mentioned corrective factors significantly modifies the quantitative results with decreases in the relative proportions of κ-CN and β-CN and conversely an increase in the relative proportions of αs-CNs, as well as β-LG.

A histogram of relative concentrations for each of the six main milk proteins measured on these 240 milk samples revealed an extreme individual variability with amplitudes from single to double, even triple, especially for κ-CN and β-LG (Supplementary material S3). Such a variability is largely explained by the main allelic variants (A and B) occurring at both these loci in the Montbéliarde population, since they are known to be expressed at different levels, as previously reported in

Table 1
Identification and quantification of the different isoforms of the six major proteins in cow's milk. Identification of bovine milk proteins found under most of the 29 peaks of the RP-HPLC chromatogram (Fig. 2) from observed molecular masses compared to theoretical masses. Quantification was achieved from the intensity of the deconvoluted mass signal.

Peak	Identification	Observed M*r* (Da)	Intensity of mass signal	Protein Family	Theoritical M*r* (Da)
peak 1	CSN3 B pyroGlu 1P 1OG (GaN, Ga, 2SA)	19953.5261	732	κ-CN B OG	19953.260
	CSN3 E pyroGlu 1P 2OG (2GaN, 2Ga, 2SA)	20319.8411	275	κ-CN E OG	20320.519
	CSN3 B pyroGlu 1P 2 OG (2GaN, 2Ga, 3SA)	20610.3938	530	κ-CN B OG	20609.844
	CSN3 B pyroGlu OP 2 OG (2GaN, 2Ga, 4SA)	20820.2702	115	κ-CN B OG	20821.118
	CSN3 B pyroGlu 1P 2 OG (2GaN, 2Ga, 4SA)	20901.5260	2315	κ-CN B OG	20901.098
	CSN3 B pyroGlu 2P 2 OG (2GaN, 2Ga, 4SA)	20980.8137	261	κ-CN B OG	20981.078
	CSN3 E pyroGlu 1P 3OG (2GaN, 2Ga, 2SA)	21268.6146	274	κ-CN E OG	21268.36
	CSN3 E pyroGlu 1P 3OG (3GaN, 3Ga, 5SA)	21558.6290	434	κ-CN E OG	21559.615
	CSN3 E pyroGlu 1P 3OG (3GaN, 3Ga, 6SA)	21850.4092	898	κ-CN E OG	21850.869
peak 2	CSN3 E pyroGlu 1P	19007.3184	8875	κ-CN E	19007.345
	CSN3 E 1P	19025.8013	475	κ-CN E	19025.365
	CSN3 E pyroGlu 2P	19088.0505	462	κ-CN E	19087.325
	CSN3 E pyroGlu 1P 1OG (GaN-Ga-SA2)	19956.3173	1644	κ-CN E OG	19955.186
peak 4	CSN3 B pyroGlu 1P 1OG (GaN, Ga, 2SA)	19953.9188	543	κ-CN B OG	19953.260
peak 5	CSN3 B pyroGlu 1P	19005.3677	13537	κ-CN B	19005.416
	CSN3 B 1P	19024.3423	755	κ-CN B	19023.436
	CSN3 B pyroGlu 2P	19085.4950	1022	κ-CN B	19085.396
	CSN3 E pyroGlu 1P 1OG (GaN-Ga-SA2)	19955.1599	1693	κ-CN E OG	19955.186
	CSN1S2 A Δexon 8 11P	24202.3812	294	αs2-CN	24202.138
peak 6	CSN1S1 B 8P ΔQ59 and Q78	23358.2451	693	αs1-CN	23358.450
	CSN3 E pyroGlu 2P 5OG (5GaN, 5Ga, 11SA)	24177.8298	650	κ-CN E OG	24117.785
	CSN1S2 A 10P	25148.2193	1901	αs2-CN	25148.348
	CSN1S2 A 11P	25228.4768	6283	αs2-CN	25228.328
	CSN3 B pyroGlu 2P 5 OG (GaN-Ga-SA)x5	25281.4687	162	κ-CN B OG	25280.874
peak 7	CSN1S2 A 12P	25308.5713	3542	αs2-CN	25308.308
	CSN1S2 A 13P	25387.6097	170	αs2-CN	25388.288
		26282.1286	380	αs2-CN ?	
peak 8	CSN1S2 A 13P	25387.1666	606	αs2-CN	25388.288
	CSN1S2 A 13P	25387.8095	606	αs2-CN	25388.288
	CSN1S2 A 14P	25467.6569	364	αs2-CN	25468.268
peak 11	CSN1S1 B 8P Δ exon 4	22069.0524	827	αs1-CN	22068.910
peak 14	CSN1S1 B 8P Δ exon 8 ΔQ59 and Q78	22426.6333	566	αs1-CN	22427.422
	CSN1S1 B 8P Δ exon 8 ΔQ59 and Q78	22427.7114	566	αs1-CN	22427.422
	CSN1S1 B 7P Δ exon 8	22602.0752	424	αs1-CN	22603.704
peak 16		20116.8349	3030	αs1-CN ?	
	CSN1S1 B 7P DHAS	23517.2581	1431	αs1-CN	23516.712
	CSN1S1 B 7P	23535.2227	3106	αs1-CN	23534.732
		23559.4071	1355	αs1-CN ?	
	CSN1S1 B 8P	23614.9064	39047	αs1-CN	23614.712
	CSN3 B pyroGlu 2P + [4OG] (4GaN, 4Ga, 5SA)	23749.9199	321	κ-CN B OG	23750.524
		23938.8005	1384		
peak 17	CSN1S1 B 8P Δ exon 12	21875.1930	2187	αs1-CN	21876.722
	CSN1S1 B 9P ΔQ59 or Q78	23565.5589	1015	αs1-CN	23566.561
	CSN1S1 B 9P	23694.8959	9497	αs1-CN	23694.692
		23715.3797	727		
		24572.9656	777		
peak 19	CSN2 B Δ exon 5 2P	22910.2747	1957	β-CN	22910.296
	CSN2 B 4P	24012.7064	2336	β-CN	24012.339
		24036.9324	1621		
		24054.4698	1473		
	CSN2 B 5P	24092.6631	53326	β-CN	24092.319
		24113.7983	2420		
		25188.6163	4777		
peak 20	CSN2 A1 Δ exon 5 2P	22841.9393	2034	β-CN	22841.187
	CSN2 A1 4P	23943.2605	2065	β-CN	23943.229
		23973.1613	1381		
	CSN2 A1 5P	24023.5263	46070	β-CN	24023.209
		24044.0983	3201		
peak 23	Gamma 3 (108-109) A1	11558.4323	318	γ3-CN A1	11558.612
peak 24	LALBA B	14186.0029	6865	α-LA	14186.064
		14207.0118	246	α-LA ?	
		14208.8616	246	α-LA ?	
peak 25	LALBA B 1P	14266.2670	218	α-LA	14266.044
	PAEP B + lactosyl	18605.3548	414	β-LG	18605.320
peak 26	PAEP B	18281.3198	22581	β-LG	18281.208
		18301.7743	1284	β-LG ?	
peak 28	PAEP A + lactosyl	18691.8373	808	β-LG	18691.300
peak 29	PAEP A	18367.3398	22976	β-LG	18367.298
		18388.2257	1319	β-LG ?	

Theoretical masses were determined using the online Expasy PeptideMass resource, as average M_r from protein sequence entries in UniProtKB. The same specific colour code defined at Fig. 1 has been adopted here for the 6 main bovine milk protein families.

other breeds (Bobe et al., 1998; Heck et al., 2008).

To validate the approach and convert the relative amounts (calculated from surfaces of the chromatogram peaks at 214 nm) into g/Kg milk, 10 milk samples from FHF cows were analyzed in LC-MS and their total protein content was determined by the amido-black technique. The average value of the total protein of the 10 milks was 31.16 g/Kg, ranging between 39 and 27.7 g/Kg. From their relative quantities expressed as a percentage of the peak area, the true amount, in g/Kg, were for each of the 6 major milk proteins: 3.00, 3.01, 10.39, 10.84, 1.05 and 2.89 g/Kg for κ-CN, αs2-CN, αs1-CN, β-CN, α-LA and β-LG, respectively, in agreement with the literature data (Supplementary material S4).

3.2.2. Quantitative determination of protein isoforms from the intensity of the mass signal

Before considering the possibility of using the Intensity of the deconvoluted Mass Signal (IMS) to quantify different isoforms of each main bovine milk protein, we first evaluated the impact of structural features, such as post-translational modifications and genetic variants, on the ionization abilities of the six main milk proteins.

3.2.2.1. Genetic variants and ionization ability. Milk samples were from cows chosen for their representativeness as genetic variants of major bovine milk proteins: variants A1, A2, A3, I and B for β-CN, variants A, E and B for κ-CN and variants A and B for β-LG. Since αs2-CN and α-LA are essentially monomorphic, these proteins were not considered. The results obtained show little or no effect of genetic variants on the ability of proteins to ionize. Indeed, the different genetic variants analyzed for both β-CN (A2, A3 and B) and κ-CN (A, B and E) gave essentially the same A_{214}/IMS ratio, except β-CN A1 which gives a slightly higher A_{214}/IMS ratio (3.8 vs 3.4), thus suggesting a lower ionization aptitude for this variant. This was confirmed by the analysis of a milk from a Montbéliarde cow, heterozygous A1/B at the *CSN2* locus which gave significantly different mass signal intensities 53,326 vs. 46,070 for variants B and A1 (peaks 19 and 20, Table 1) whereas A_{214} were very similar for these two peaks. It is worth noting that the A_{214}/IMS ratio remained of the same order for caseins, while it is about twice as low for variants of β-LG, suggesting a better ionization efficiency for this protein.

3.2.2.2. Effect of the phosphorylation level of caseins. Regarding αs1-CN and β-CN, little or no effect of phosphorylation was observed (results not shown). By contrast, a significant effect was recorded for αs2-CN that displays the highest phosphorylation level, ranging between 7 and 14 phosphate groups (Fang et al., 2016). Consequently, the risk of underestimation of αs2-CN in the case of absolute quantification is real, but not in the case of a relative quantification (comparison between milk samples). The effect was even more marked when the injected volume of the sample was low (Supplementary material S6). With an injected volume of 5 μL the impact on the other caseins (αs1-CN and β-CN) is no longer negligible. However, it must be considered, at least in cattle, that the phosphorylation level of those caseins is rarely zero, which further gives a relative dimension to the observed effect and justifies not taking into account this factor, except for αs2-CN.

3.2.2.3. Quantitative determination of protein isoforms present in each peak. After having analyzed the effects of the main influencing factors (genetic variants and PTM status), correction factors were determined and applied to convert the IMSs (deconvoluted spectra) in relative proportion (expressed in %) of each identified protein molecule. Thus, it is possible to decompose each compound corresponding to the different UV peaks into their different isoforms and quantify them. For example, the B and A1 variants of β-CN, which correspond to peaks 19 and 20 in Fig. 2, respectively, can be broken down in isoforms according to their phosphorylation levels (4 and 5P). Those peaks also contained molecular species very likely derived from β-CN that we were not able to identify, based on their molecular masses, as well as a splicing variant (exon 5 skipping) with two phosphate residues (Table 1). The IMSs observed with these different isoforms makes it possible to quantify each of them. It appears that the 5P isoform, that is overwhelmingly predominant, represents nearly 80% of all isoforms both in peaks 19 (B variant) and 20 (A1 variant). The 4P and exon five skipped isoforms, each represent only *ca.* 3% of all β-CN isoforms. Under the optimized chromatographic conditions used, γ3-CN is eluted after β-CN A3 (peak 26, Fig. 1; peak 23 in Fig. 2 and Table 1), and can therefore be quantified, in contrast to what has been previously reported (Bobe et al., 1998; Groen et al., 1994).

Similarly, multiple α_{s1}-CN isoforms, including splicing isoforms arising from skipping of exons 4, 8 or 12, from the usage of intra-exon cryptic splice sites (ΔQ78 and/or ΔQ59), as well as isoforms with different phosphorylation levels, which are distributed in peaks 11, 14, 16 and 17 (Fig. 2), can be individually quantified from their IMS (Table 1). Again, isoforms corresponding to the full-length protein, with eight (peak 16) and nine (peak 17) phosphate groups that account for 60 and 16% of the αs1-CN family, respectively, are mainly represented, whereas splicing variants account for *ca.* 10%. Furthermore, it is worth noting the possible presence of small amounts, in peak 16, of a glycosylated isoform of κ-CN B that can be reassigned to the κ-CN family. A similar situation was observed for peak 6, which contained, in addition to αs2-CN 10 and 11P, a molecule possibly corresponding to a glycosylated isoform of κ-CN B-2P.

αs2-CN is distributed in 3 peaks (6, 7 and 8), not completely resolved, corresponding to differently phosphorylated isoforms (10 to 14P) that can be easily quantified individually from the IMS data, whereas quantification based on peak area would be impossible. Thus, we were able to estimate the proportion of each αs2-CN phosphorylation isoform, as a fraction of total αs2-CN. αs2-CN with 11 and 12P were the most abundant isoforms with 47 and 26%, respectively, in agreement with previously reported results in the Montbéliarde breed (Fang et al., 2016). However, the relative proportions of these isoforms are prone to vary widely among individual cows. Given the involvement of phosphate in the casein micelle structure, there is a range of genetic progress to explore in this direction.

Although they are rather well resolved and possibly quantified from absorbance at 214 nm, αs1-CN B isoforms with 8 and 9 phosphate were more precisely quantified using the IMS, as shown in Fig. 3. Interestingly, by slightly modifying the chromatographic conditions, exemplified by milk #57 protein profiling (Fig. 3), we succeeded in improving the separation between αs1-CN isoforms (peaks 22 to 26), including genetic variants B and C, which allows their quantification as well as a more accurate quantification of αs1-CN-8P and αs1-CN-9P (Fig. 3, Supplementary material S5). However, since additional isoforms arising mainly from splicing anomalies and a lower phosphorylation isoform (αs1-CN-7P) also occur in several peaks (from peak 18 to 26) together with unidentified masses, mass signal intensity is undoubtedly the best way to quantify precisely if not all of the αs1-CN isoforms, at least a large number of them. In such a way, we estimated the relative proportions of αs1-CN with 7P, 8P and 9P to be *ca.* 5, 70 and 25%, respectively. For comparison, the ratio of αs1-CN 8P to αs1-CN 9P based on the area of peaks (23 and 25) at 214 nm is 3.5, vs. 2.8 in IMS. In addition, in such chromatographic conditions, we also achieved the beginning of a resolution of β-CN A2 and I variants (Fig. 3, peaks 34 and 35).

A thorough analysis of the LC–MS data (Supplementary material S5 and Fig. 3) from the LC-MS in long gradient conditions revealed that:

1 – κ-casein (peaks 1–9) exists, whether variant A or B, mainly as a mono-phosphorylated form, together with a minor isoform (< 10%) bearing two phosphate residues, the main phosphorylation site being residue S149, whereas the second one is S121, according to Holland and Boland (2014);

2 – in agreement with literature (Van Eenennaam & Medrano,

Table 2
Average distribution of the different protein families, estimated on a panel of 240 individual milk samples from Montbéliarde cows, and compared with data from the literature.

	Current study (UV 214 nm without correction)	Current study (UV 214 nm with correction)	Current study (UV 214 nm with correction and reassignment of proteolysis products)	Bobe et al. (1998)	Fang et al. (2016)	Heck et al. (2008)	Jensen et al. (2012) *
κ-CN	10.13	8.81	9.10	16.9	9.03	8.4	10.85
αs2-CN	8.15	8.45	8.59	8,0	4.41	10.1	5.35
αs1-CN	28.67	34.32	34.95	32.2	32.92	33.6	24.66
β-CN	33.40	29.01	31.57	28.6	28.14	27.2	34.36
Total CN	80.35	80.59	84.21	85.5	74.5	79.3	83.23
α-LA	2.58	2.62	2.70	3.8**	3.54	2.4	3.18
β-LG	7.79	10.91	10.98	10.5	12.16	8.3	6.51
Total 6 main milk proteins	90.72	94.12	97.90	99.8	90.2	90,0	92.92
Others	9.28	5.88	2.10	0.2	9.8	10,0	7.08

*Holstein, good milk coagulation class; total CN = area of all peaks in the CN-elution interval of the chromatogram.
**Including BSA that co-elutes with α-LA Correction was for absorbance at 280 nm. Re-assignment of proteolysis product was performed considering that others when > 5% were mainly proteolysis products from caseins: β-CN (70%), αs1-CN (17%) and κ-CN (13%).

1991), the B variant is more expressed (+80%) than the A;
3 – variant B shows a greater diversity of glycan patterns and a higher level of glycosylation (number of modified sites), the major glycan motif being a tetrasaccharide composed of Galactose (Ga), N-acetylgalactosamine (GaN) and Sialic or neuraminic Acid (SA) of the form SA-Ga-GaN[SA];
4 – the glycoforms mainly represented in variants A and B are different, 1P-1OG (1 tetrasaccharide unit) on variant A and 1P-2OG (2 tetrasaccharide units) on variant B;
5 – taking into account the impact of glycosylation on the IMS (Wada, 2012), the Glyco/Non-Glyco ratio was estimated to be 50/50 for variant A while it was close to 60/40 for variant B, consistent with the literature (Holland & Boland, 2014).

4. Qualification of the LC-MS method

The global validation of the RP-HPLC method applied to bovine milk has been previously demonstrated for chromatograms recorded at 214 nm (Bonfatti et al., 2008; Bordin et al., 2001). Bonfatti et al. (2008) reported a linear relationship between the concentrations of milk proteins and peak areas over the range of bovine milk protein content. They estimated the precision (repeatability and reproducibility) as satisfactory for both retention times and peak areas.

In agreement with their results, we also observed a good reproducibility and repeatability for retention times (even from one column to another) and peak areas (results not shown), provided corrective factors, taking into account the bias introduced by specific absorbance at 280 nm, have been determined for each protein and applied to the measurement of absorbance at 214 nm.

However, it remains to validate the quantitative determination of protein isoforms from the deconvoluted mass signals by evaluating the linearity and the intermediate precision (repeatability and reproducibility) as well as the accuracy of the LC–MS method.

4.1. Results of linearity tests

Each sample corresponding to the 11 levels of milk protein concentration was analyzed by the LC-MS method, in duplicate. A simple linear regression and a curvilinear regression (order 2) were calculated systematically by taking the raw deconvoluted IMS data on the ordinates and the theoretical (calculated) milk protein levels on the abscissa. In case of non-linearity, a second treatment was performed as a linear regression on the graphically observed linearity range. An Ar/At ratio (Ar and At being the amplitude of IMS residues and amplitude of IMS units, respectively) was calculated on each linear regression performed (over the entire rate range and the smaller range). For the six proteins tested, the response is not linear over the entire range (from 0 to 87 $g.L^{-1}$ of total protein), *i.e.* between 0 and 5.2 $g.L^{-1}$ of non-glycosylated κ-CN, for which we observe a ratio Ar/At = 15.9%. Interestingly, the response of the method was curvilinear over the entire range (Fig. 4A). The R^2 increased from 0.9769 for linear regression to 0.9929 for curvilinear regression. However, a linear response was observed between 0 and 43 $g.L^{-1}$ total protein (*i.e.* from 0 to 2.6 $g.L^{-1}$ non-glycosylated κ-CN casein). The linear regression performed over this range gave an Ar/At ratio of 3.3% and an R^2 of 0.9966. Quite the same situation was recorded with all four caseins, as well as with whey proteins (Supplementary material S6). Interestingly, the widest linear range registered was between 8 and 78 $g.L^{-1}$ of total milk protein, with α-LA (0.27 to 2.48 $g.L^{-1}$). A linear regression over this range gave an Ar/At ratio of 6.7%.

From an overall point of view, we conclude that the method is not strictly linear over the entire range tested (0 to 87 $g.L^{-1}$ protein). However, the method, as parameterized, is linear over a total protein content range of 8 to 43 $g.L^{-1}$, which is more than sufficient for analyzing cow and goat milks. For the analysis of ewe's milk, a ½ dilution of the starting sample is required.

4.2. Intermediate precision

The intermediate precision of the LC-MS method was evaluated by estimating the intra-laboratory repeatability and reproducibility. The results obtained for the six major bovine milk proteins and some isoforms and variants are summarized in Supplementary material S6. The Relative Standard Deviation for repeatability (RSDr) values varied between 2.2 and 5.3% according to the protein (except for α-LA and αs2-CN A-12P for which the RSDr values were 11.7 and 13.6%, respectively), genetic variants and isoforms tested.

The following parameters, Sr and SR (standard deviation of repeatability and reproducibility within the laboratory) and r and R (maximum deviation between doubles and maximum reproducibility deviation), were determined for each type of expression, *i.e.* in relative percentage of total protein (%) and in g per liter of milk ($g.L^{-1}$), for

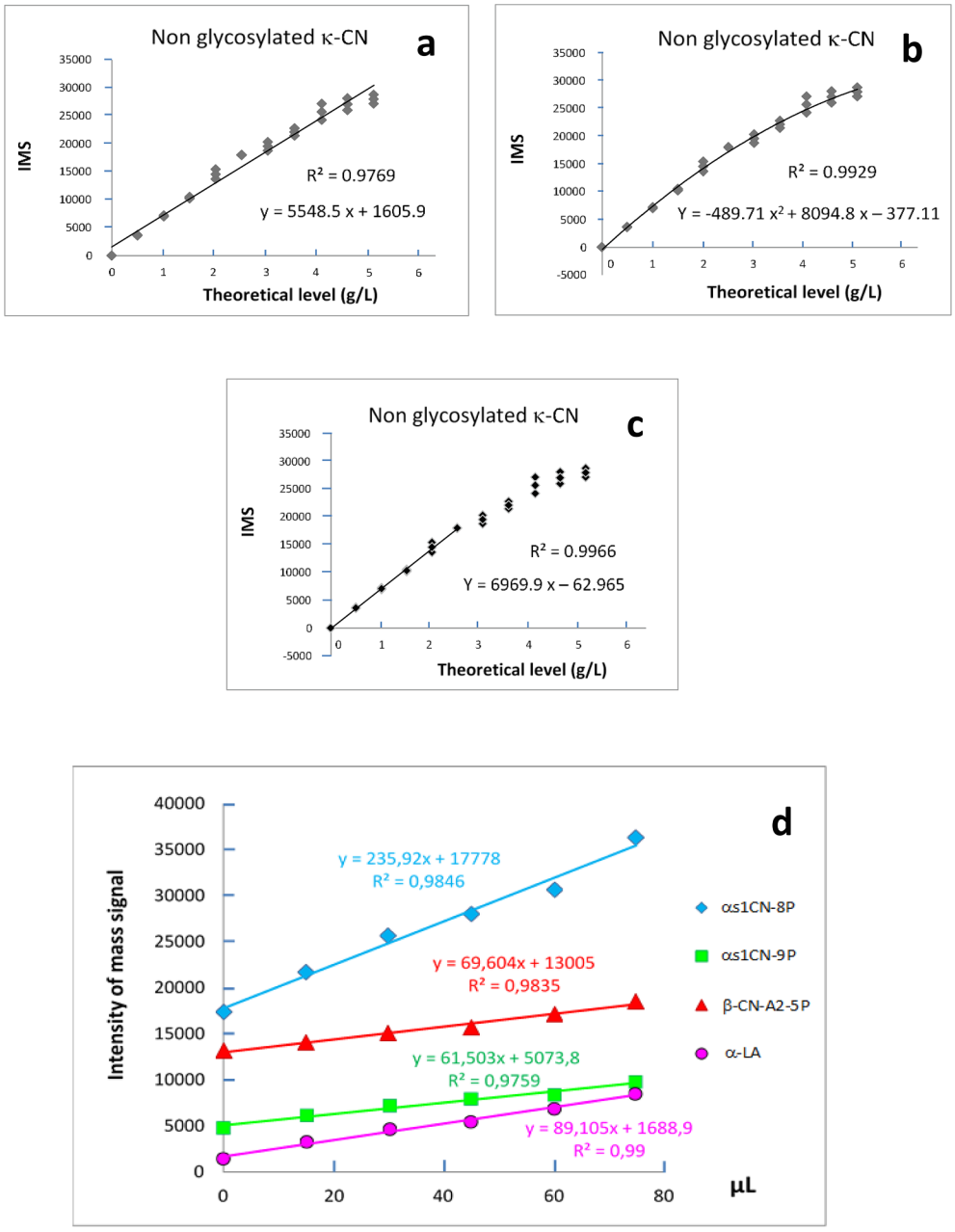


Fig. 4. Qualification of the LC-MS method to quantify allelic variants as well as PTM isoforms of the six main bovine milk proteins from the intensity of mass signal. Linear (a) and curvilinear (b) tests on non-glycosylated κ-CN. Linear test on the graphically observed linearity range (c). Accuracy test (d) was estimated from increasing spiking amounts of three purified (purity ranging between 90 and 95%) bovine milk proteins: αs1-CN B-(8P + 9P), β-CN A2-5P and α-LA of which concentrations were 0.725, 0.118 and 0.157 µg/µL, respectively.

each milk protein. The results obtained for the six major bovine milk proteins and some isoforms and variants are summarized in Supplementary material S6. The Relative Standard Deviation for reproducibility (RSDR) values ranged between 3.68 and 16.92% according to the proteins, genetic variants and isoforms tested.

4.3. Accuracy

From samples spiked at six different levels (see M & M section) with pure protein isolates, Average Recovery Coefficients (ARC) have been calculated for αs1-CN-8P and 9P, β-CN A2-5P and α-LA (Fig. 4). ARC ranged between 94.3% (αs1-CN-8P) and 101.4% (α-LA), with ARC of 97.0 and 100.6% for αs1-CN-9P and β-CN A2-5P, respectively (Supplementary material S6).

5. Concluding remarks

Originally, this method was designed and developed as an anchoring method for the establishment of equations to predict the composition of six major milk proteins (4 caseins, β-LG and α-LA) from MIR spectra in the bovine milk protein fraction (Ferrand et al., 2012). However, in view of the power and potential of the tool, we continued its development to provide even more precise and detailed information, leading to the identification of most of the isoforms of each protein family including genetic variants, splicing variants and isoforms resulting from post-translational modifications. To achieve this, the chromatographic conditions were optimized and the theoretical mass database was concomitantly gradually expanded to reach today nearly 3000 molecular mass references. The method still needs to be

improved, even if it has already led to significant progress in various fields, particularly in Genetics (Sanchez et al., 2016) and Physiology (Fang et al., 2016). An issue that remains to be addressed concerns the identity of minor compounds present under a peak, displaying masses not listed in our database. Those are often close to the theoretical mass of the main compound, but likely result from chemical modifications (adducts) occurring during sample storage and/or processing. This is particularly the case for β-LG and α-LA that occur with multiple masses. Nevertheless, depending on the status of the cows at each locus (homozygous or heterozygous), up to one hundred molecules derived from the six major milk proteins can be identified and quantified in an individual milk sample, even though ca. 1/3 of the masses giving a signal of significant intensity (globally accounting for less than 6% of the milk proteins) could not be formally identified.

Given its "versatility", this profiling method has no equivalent. Indeed, besides a coarse composition, by protein family, it makes a fine phenotyping possible, which goes well beyond the simple identification of genetic variants. Indeed, its accuracy and its resolving power provide the possibility to identify and quantify most of the protein isoforms arising from PTM, from defective splicing as well as proteolysis products (with the possibility of reassigning them to the family of origin).

It is important to stress that the pre-treatment of the sample is an essential point in order to obtain reliable results. The option taken was to make this pre-treatment minimal. The samples therefore underwent a simple centrifugation skimming process, under mild conditions, so that the largest micelles do not precipitate. In such a way, the differences between observed and theoretical masses (accuracy) are very small (lower than 0.35 Dalton on average) for the main isoforms, which makes it easier to interpret the data. However, identification based on the observed molecular mass, remains a relative limitation because in rare circumstances this feature alone is not sufficient. Indeed, in some cases it is difficult to conclude definitively on the nature of the "identified" molecule. That is why it is highly recommended to take into account the elution time even though it is not always completely consistent: e.g. αs1-CN B 8P Δ (Q59 and Q78), which was eluted in peak 6 (Table 1) and consequently well upstream of the elution zone of αs1-CN. Similarly, in the same milk sample, a mass of 22,691.36 Da identified a highly glycosylated κ-CN with 2P groups (theoretical mass 22,691.54 Da) which was eluted in peak 26 among αs1-CN isoforms. It is therefore useful to be able to verify the nature of the molecule, which LC-MS/MS could allow.

Another crucial point is the time spent for an analysis. While the run itself, set at 30 min. for routine analyses, remains acceptable, the interpretation and extraction of data is by far longer and requires the development and implementation of automation tools (in progress).

It is also valuable to highlight here the wide range of investigation possibilities in the dairy industry. This method has proven effective in monitoring the degradation of caseins (αs1- and β-CN) during their conservation (milk bank kept at the Centre de Ressources Biologiques (CRB) animales, INRAE, Jouy). This ability to precisely identify and quantify large proteolysis products is a key tool for monitoring casein degradation during cheese ripening (under study). It has also demonstrated its effectiveness in assessing the impact of different farming practices (dietary restrictions, single milking) on the protein composition of milk (unpublished results), and it could potentially be useful to characterize ingredients used to prepare infant formulas.

Finally, this method can be applied to all mammalian species with a minimum of RP-HPLC separation condition development and the creation of a specific mass database. In this respect, recent studies on small ruminants (goats, "MilkChEST" program, Martin et al., manuscript in preparation), rabbits (maternal high-fat/high-sugar obesogenic diet, Hue-Beauvais et al., 2017) and camelids (Ryskaliyeva et al., 2019) are worth mentioning.

Author contributions

G.M. designed, performed and supervised the experiments, participated in data interpretation, and contributed to the building of the bovine theoretical mass database; L.B. participated in LC-MS analyses and in the development of the method, contributed to the building of the bovine theoretical mass database; Z.K. performed LC-MS analyses and contributed to the development of the method; Ph.T. supervised experiments performed and calculation for the method validation and qualification; P.M. conceived and supervised the research, participate to data interpretation and contributed to the building of the bovine theoretical mass database. The manuscript was written by P.M., revised and approved by G.M. and L.B. All the authors reviewed and approved to the final manuscript.

Declaration of Competing Interest

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

Acknowledgements

The authors would like to warmly thank the Centre National Interprofessionnel de l'Economie Laitière (CNIEL), APIS-GENE, the Ministère de l'Agriculture & de l'Alimentation (CASDAR) and the Agence Nationale pour la Recherche (ANR) for funding the following research programs: "PhénoFinLait", "LactoScan", Mass Quant Milk" and "FROM'MIR". In addition, we would also like to thank animal staffs of Le Domaine expérimental INRAE du Pin-au-Haras and La ferme expérimentale INRAE de Mirecourt, Le Conseil en Elevage du Doubs & du Territoire de Belfort (CEL 25-90) as well as the Institut de l'Elevage, for milk sampling. We thank Céline Henry and Lydie Oliveira-Correia from the INRAE Proteomics core facility PAPPSO, whose incomparable expertise is only matching with their willingness, for performing LC-MS/MS analyses. Finally, we would like to thank Wendy Brand-Williams for editing the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.fochx.2020.100080.

References

Amalfitano, N., Cipolat-Gotet, C., Cecchinato, A., Malacarne, M., Summer, A., & Bittante, G. (2019). Milk protein fractions strongly affect the patterns of coagulation, curd firming, and syneresis. *Journal of Dairy Science, 102*(4), 2903–2917. https://doi.org/10.3168/jds.2018-15524.

Bobe, G., Beitz, D. C., Freeman, A. E., & Lindberg, G. L. (1998). Separation and quantification of bovine milk proteins by reversed-phase high-performance liquid chromatography. *Journal of Agricultural and Food Chemistry, 46*, 458–463.

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

Bordin, G., Cordeiro Raposo, F., De La Calle, B., & Rodriguez, A. R. (2001). Identification and quantification of major bovine milk proteins by liquid chromatography. *Journal of Chromatography A, 928*, 63–76.

Boutrou, R., Gaudichon, C., Dupont, D., Jardin, J., Airinei, G., Marsset-Baglieri, A., ... Léonil, J. (2013). Sequential release of milk protein-derived bioactive peptides in the jejunum in healthy humans. *The American Journal of Clinical Nutrition, 97*, 1314–1323.

Fang, Z. H., Visker, M. H. P. W., Miranda, G., Delacroix-Buchet, A., Bovenhuis, H., & Martin, P. (2016). The relationships among bovine α_s-casein phosphorylation isoforms suggest different phosphorylation pathways. *Journal of Dairy Science, 99*, 8168–8177.

Ferrand M., Miranda G., Guisnel S., Larroque H., Leray O., Lahalle F., Brochard M., &

Martin P. (2012). Determination of protein composition in milk by mid-infrared spectrometry. ICAR Meeting, 28 mai-1er juin, Cork, Irlande.
Frederiksen, P. D., Andersen, K. K., Hammershøj, M., Poulsen, H. D., Sørensen, J., Bakman, M., ... Larsen, L. B. (2011). Composition and effect of blending of non-coagulating, poorly coagulating, and well-coagulating bovine milk from individual danish holstein cows. *Journal of Dairy Science, 94*, 4787–4799.
Groen, A. F., van der Vegt, R., van Boekel, M. A. J. S., de Rouw, O. L. A. M., & Vos, H. (1994). Case study on individual animal variation in milk protein composition as estimated by high-pressure liquid chromatography. *Netherlands Milk and Dairy Journal, 48*, 201–212.
Heck, J. M. L., Olieman, C., Schennink, A., van Valenberg, H. J. F., Visker, M. H. P. W., Meuldijk, R. C. R., & van Hooijdonk, A. C. M. (2008). Estimation of variation in concentration, phosphorylation and genetic polymorphism of milk proteins using capillary zone electrophoresis. *International Dairy Journal, 18*, 548–555.
Holland, J. W., & Boland, M. J. (2014). Post-translational modifications of caseins. In H. Singh, M. Boland, & A. Thompson (Eds.). *Milk proteins, from expression to food* (pp. 141–168). (2nd ed.). Amsterdam: Elsevier.
Hue-Beauvais, C., Miranda, G., Aujean, E., Jaffrezic, F., Devinoy, E., Martin, P., & Charlier, M. (2017). Diet-induced modifications to milk composition have long-term effects on offspring growth in rabbits. *Journal of Animal Science, 95*, 761–770.
Ikonen, T., Ahlfors, K., Kempe, R., Ojala, M., & Ruottinen, O. (1999). Genetic parameters for the milk coagulation properties and prevalence of noncoagulating milk in finnish dairy cows. *Journal of Dairy Science, 82*, 205–214.
Jann, O., Ibeagha-Awemu, E. M., Ozbeyaz, C., Zaragoza, P., Williams, J. L., Ajmone-Marsan, P., ... Erhardt, G. (2004). Geographic distribution of haplotype diversity at the bovine casein locus. *Genetics, Selection, Evolution, 36*, 243–257.
Jaubert, A., & Martin, P. (1992). Reverse-phase HPLC analysis of goat caseins. Identification of αS1 and αS2 genetic variants. *Le Lait, 72*, 235–247.
Jensen, H. B., Holland, J. W., Poulsen, N. A., & Larsen, L. B. (2012). Milk protein genetic variants and isoforms identified in bovine milk representing extremes in coagulation properties. *Journal of Dairy Science, 95*, 2891–2903.
Joudu, I., Henno, M., Värv, S., Kaart, T., & Kärt, O. (2007). Milk protein genotypes and milk coagulation properties of Estonian native cattle. *Agricultural and Food Science, 16*, 222–231.
Korhonen, H., & Pihlanto, A. (2007). Technological options for the production of health-promoting proteins and peptides derived from milk and colostrum. *Current Pharmaceutical Design, 13*, 829–843.
Léonil, J., Molle, D., Gaucheron, F., Arpino, P., Guenot, P., & Maubois, J.-L. (1995). Analysis of major bovine-milk protein by online high-performance liquid-chromatography and electrospray-ionization mass-spectrometry. *Lait, 75*, 193–210.
Mamone, G., Caira, S., Garro, G., Nicolai, A., Ferranti, P., Picariello, G., ... Addeo, F. (2003). Casein phosphoproteome: Identification of phosphoproteins by combined mass spectrometry and two-dimensional gel electrophoresis. *Electrophoresis, 24*, 2824–2837.
Martin, P., Bianchi, L., Cebo, C. & Miranda, G. (2013). Genetic polymorphism of milk proteins. In McSweeney, P. L. H., & Fox, P. F. (Eds.). Advanced dairy chemistry: volume 1A: Proteins: Basic Aspects, (4th ed. pp. 463-514). Springer Science +Business Media New York.
Martin, P., Cebo, C. & Miranda, G. (2013). Interspecies comparison of milk proteins: quantitative variability and molecular diversity. In McSweeney, P. L. H., & Fox, P. F. (Eds.). Advanced dairy chemistry: volume 1A: Proteins: Basic Aspects, (4th ed. pp. 387-429). Springer Science+Business Media New York.
Meisel, H. (2004). Multifunctional peptides encrypted in milk proteins. *BioFactors, 21*, 55–61.
Miralles, B., Leaver, J., Ramos, M., & Amigo, L. (2003). Mass mapping analysis as a tool for the identification of genetic variants of bovine β-casein. *Journal of Chromatography A, 1007*, 47–53.
Miralles, B., Rothbauer, V., Manso, M. A., Amigo, L., Krause, I., & Ramos, M. (2001). Improved method for the simultaneous determination of whey proteins, caseins and para-κ-casein in milk and dairy products by capillary electrophoresis. *Journal of Chromatography A, 915*, 225–230.
Miranda, G., Krupova, Z., Bianchi, L., & Martin, P. (2013). A novel LC-MS protein profiling method to characterize and quantify individual milk proteins and multiple isoforms. In 10th annual International Milk Genomics Consortium symposium. October 1-3, 2013 U.C. Davis Conference Center, Davis, California USA.
Miranda, G., Mahé, M.-F., Leroux, C., & Martin, P. (2004). Proteomic tools to characterize the protein fraction of equidae milk. *Proteomics, 4*, 2496–2509.
Mohanty, D. P., Mohapatra, S., Misra, S., & Sahu, P. S. (2016). Milk derived bioactive peptides and their impact on human health – A review. *Saudi Journal of Biological Sciences, 23*, 577–583.
Nieuwenhuijse, J. A., van Boekel, M. A. J. S., & Walstra, P. (1991). On the heat-induced association and dissociation of proteins in concentrated skim milk. *Netherlands Milk and Dairy Journal, 45*, 3–22.
Ortega, N., Albillos, S. M., & Busto, M. D. (2003). Application of factorial design and response surface methodology to the analysis of bovine caseins by capillary zone electrophoresis. *Food Control, 14*, 307–315.
Otte, J., Zakora, M., Kristiansen, K. R., & Qvist, K. B. (1997). Analysis of bovine caseins and primary hydrolysis products in cheese by capillary zone electrophoresis. *Le Lait, 77*, 241–257.
Recio, I., Pérez-Rodríguez, M., Ramos, M., & Amigo, L. (1997). Capillary electrophoretic analysis of genetic variants of milk proteins from different species. *Journal of Chromatography A, 768*, 47–56.
Rijnkels, M. (2002). Multispecies comparison of the casein gene loci and evolution of casein gene family. *Journal of Mammary Gland Biology and Neoplasia, 7*(3), 327–345.
Ryskaliyeva, A., Henry, C., Miranda, G., Faye, B., Konuspayeva, G., & Martin, P. (2019). Alternative splicing events expand molecular diversity of camel CSN1S2 increasing its ability to generate potentially bioactive peptides. *Scientific Reports, 9*, 5243.
Sanchez, M. P., Govignon-Gion, A., Ferrand, M., Gelé, M., Pourchet, D., Amigues, Y., ... Boichard, D. (2016). Whole-genome scan to detect quantitative trait loci associated with milk protein composition in 3 French dairy cattle breeds. *Journal of Dairy Science, 99*, 8203–8215.
Van Eenennaam, A. L., & Medrano, J. F. (1991). Differences in allelic protein expression in the milk of heterozygous K-casein cows. *Journal of Dairy Science, 74*, 1491–1496.
Vincent, D., Elkins, A., Condina, M. R., Ezernieks, V., & Rochfort, S. (2016). Quantitation and identification of intact major milk proteins for high-throughput LC-ESI-Q-TOF MS analyses. *PLoS ONE, 11*(10), e0163471.
Visker, M. H. P. W., Dibbits, B. W., Kinders, S. M., van Valenberg, H. J. F., van Arendonk, J. A. M., & Bovenhuis, H. (2011). Association of bovine beta-casein protein variant i with milk production and milk protein composition. *Animal Genetics, 42*, 212–218.
Visser, S., Slangen, C. J., & Rollema, H. S. (1991). Phenotyping of bovine milk proteins by reversed-phase high-performance liquid chromatography. *Journal of Chromatography A, 548*, 361–370.
Wada, Y. (2012). Label-free analysis of o-glycosylation site-occupancy based on the signal intensity of glycopeptide/peptide ions. *Mass Spectrometry (Japan), 1*(2), A0008.
Wedholm, A., Larsen, L. B., Lindmark-Månsson, H., Karlsson, A. H., & Andrén, A. (2006). Effect of protein composition on the cheese-making properties of milk from individual dairy cows. *Journal of Dairy Science, 89*, 3296–3305.