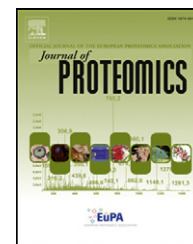Available online at www.sciencedirect.com

SciVerse ScienceDirect

www.elsevier.com/locate/jprot

A peptidomic approach to biomarker discovery for bovine mastitis

Rozaihan Mansor[a,1], William Mullen[b], Amaya Albalat[b], Panagiotis Zerefos[b,c],
Harald Mischak[b,d], David C. Barrett[e,2], Andrew Biggs[f], P. David Eckersall[a,*]

[a]Institute of Infection, Immunity and Inflammation, University of Glasgow, Glasgow, UK
[b]Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, UK
[c]Biotechnology Division, Biomedical Research Foundation, Academy of Athens, Athens, Greece
[d]Mosaique Diagnostics, Hannover, Germany
[e]School of Veterinary Medicine, University of Glasgow, Glasgow, UK
[f]Vale Veterinary Laboratories, Devon, UK

ARTICLE INFO

ABSTRACT

Bovine mastitis is usually caused by either Gram positive or Gram negative bacteria, reducing the quantity and quality of milk produced. This investigation using capillary electrophoresis and mass spectroscopy, studied peptides in milk from cows with clinical mastitis in comparison to milk from healthy cows to identify biomarkers for mastitis. In addition, the milk peptidome from udders infected with Gram positive Staphylococcus aureus (S. aureus) or with Gram negative Escherichia coli (E. coli), was examined to assess differential diagnosis between the causative agent. Comparison of the peptidome between healthy (n = 10) and mastitic milk (n = 27) identified 154 peptides for a biomarker panel which in a model for diagnosis of mastitis showed 100% sensitivity and specificity. β-casein and α_{s1} casein provided the majority of peptides identified in this model. The peptidome comparison of milk from mastitis cases caused by S. aureus (n = 8) or E. coli (n = 11) revealed a biomarker panel of 47 peptides which discriminated between cause of infection with a sensitivity of 75% and a specificity of 100%. β-casein fragments were the most common of the peptides in this model. Peptide biomarkers of milk could be used in the diagnosis of mastitis and can discriminate between these two bacterial causes.

Biological significance
The paper describes an innovative approach to the use of gel free proteomics to identify the peptides that are present in milk during clinical mastitis, which is a major cause of loss of production to dairy farmers worldwide. The use of capillary electrophoresis, liquid chromatography and mass spectrometry has been able to identify panels of peptides which can be used for disease diagnosis and for differential diagnosis of the causative bacteria of the infections of the mammary gland. As well as contributing to our knowledge of the

* Corresponding author at: Institute of Infection, Immunity and Inflammation, Garscube Estate, University of Glasgow, Bearsden Rd, Glasgow G61 1QH, UK. Tel.: +44 141 330 5732; fax: +44 141 330 2483.
E-mail address: david.eckersall@glasgow.ac.uk (P.D. Eckersall).
[1] Now at: Universiti Putra Malaysia,Selangor, Malaysia.
[2] Now at School of Veterinary Sciences, University of Bristol, Langford House, Langford, Bristol, UK.

pathophysiology of bovine mastitis the results could be the basis of improved detection and differential diagnosis of the disease.

1.   Introduction

Mastitis is an inflammatory disease of the mammary gland parenchyma which is characterized by a range of physical and chemical changes of milk and pathological changes in the udder tissues [1,2]. Significant milk changes that can be observed in bovine mastitis are the presence of clots in milk, milk discolouration and high levels of leukocyte numbers in affected milk which lead to a rise in somatic cell count (SCC). Furthermore, clinical signs occur in bovine mastitis including swelling, heat and pain in the udder. Mastitis is usually caused by bacterial infection with major pathogens being *Streptococcus agalactiae*, *Staphylococcus aureus* (*S. aureus*) and *Mycoplasma bovis* as well as environmental pathogens, including *Streptococcus* species (*Streptococcus uberis* and *Streptococcus dysgalactiae*) and environmental coliforms Gram negative bacteria *Escherichia coli* (*E. coli*), *Klebsiella* spp., *Citrobacter* spp., *Enterobacter* spp., *Enterobacter faecalis* and *Enterobacter faecium*, and other Gram negative bacteria such as *Serratia*, *Pseudomonas* and *Proteus* [1,3,4].

Mastitis is one of the most prevalent diseases in dairy cows. It is an endemic disease and is the most frequent and most costly disease affecting dairy herds worldwide [5–7]. It affects the quality of the milk through changes in milk composition and has marked economic consequences for the dairy industry. The economic losses due to mastitis can be caused by milk production losses, use of treatments, discarded milk, veterinary services, labour, product quality, materials and investments, diagnostics, other diseases and culling [5]. Persistent decrease in milk production is the main detrimental effect that contributes to the economic impact of mastitis [8] and can be caused by both clinical and subclinical mastitis. While identification of clinical mastitis can be achieved by identification of the clinical signs described above, subclinical mastitis where these signs are not evident is a more problematic diagnostic task. Currently subclinical mastitis is identified by measurement of SCC in milk [9] which required submission of sample to a laboratory for automated cell counting or associated measures such as conductivity or lactate dehydrogenase activity. The California Mastitis test which is a measure of SCC by agglutination may also be applied on farm [7]. Identification of alternative biomarkers of mastitis which could be adapted to rapid, on-farm diagnostic systems would be a valuable tool for early detection and treatment of mastitis.

Changes are known to take place in the milk proteome during mastitis. Clinical mastitis causes decreases in caseins, α-lactalbumin and β-lactoglobulin [10] and increases in serum albumin and serotransferrin. Further studies have identified additional proteins with altered expression in milk from dairy cows with experimental induction of mastitis [11–13]. These studies have revealed increased concentrations of chaperonins, DNA-binding proteins, enzymes, transport proteins and structural proteins, complement factors C3 and C4, lactoferrin, transferrin, apolipoprotein AI, fibrinogen, glycosylation-dependent cell adhesion molecule-1, peptidoglycan recognition receptor protein and cyclic dodecapeptide-1.

In addition proteins secreted from neutrophils and lymphocytes have been identified.

An investigation by Ibeagha-Awemu et al. [14] has suggested that different pathogens have varied effects on the milk proteome. Thus, in natural cases of the disease, the proteome of whey from cows with mastitis caused by *E. coli* and *S. aureus* differed substantially in comparison to whey from healthy cows as revealed by one-dimensional, liquid chromatography tandem MS (LC–MS/MS). A total of 73 proteins were found to be significantly different between normal whey and whey from quarters infected with either *E. coli* or *S. aureus*. Infection with *E. coli* caused a greater change in the expression of the proteins.

A potential cause of alteration of the milk proteome in mastitis is proteolysis of milk protein by bacterial or endogenous proteases. Following induction of mastitis by infusion of lipoteichoic acid (LTA), the toxin of *S. aureus*, Larsen et al. [15] identified, by 2-D gel electrophoresis, hydrolysis of β- and α_{S1}-caseins 6 h after infusion with the toxin. Examination of the peptides at this and later time points, using reversed-phase HPLC and MALDI-TOF MS/MS, detected increases in the milk peptidome with several different peptides being characterized and found to originate from α_{S1}- and β-caseins.

The analysis of change in the peptide content of biological fluid such as serum or urine is increasingly being used as a means to discover biomarkers for the diagnosis and monitoring of disease. There have been a number of studies, especially using urine, that have developed biomarkers models for a number of different diseases using capillary electrophoresis (CE) and mass spectrometry (MS). Those investigated so far range from renal disorders [16,17] to cardiovascular disease [18] and diabetes [19]. Many of these biomarker profiles have been recorded in a database made up from over 13,000 samples [20]. These biomarkers are not based on a single analyte but are composed of a group or panel of proteins/peptides. Changes in the profile or pattern can indicate a trend toward or away from the disease state.

It is likely that changing patterns of the peptides of milk could provide a source of biomarkers for mastitis so that application of advanced peptide biomarker detection and analysis could be a valuable addition tool in the diagnosis of mastitis.

The aim of this investigation was to undertake a study of the peptides present in milk from healthy dairy cows in comparison to milk from cows with naturally occurring clinical mastitis to determine if peptide biomarkers in milk could be identified for use in disease diagnosis. A second aim was to determine if peptide biomarkers could be identified which could distinguish between the bacteria causing naturally occurring cases of mastitis by analysis of samples in which either *E. coli* or *S. aureus* had been identified as the pathogenic agent as a means for differential diagnosis between bacterial cause [4].

2.   Material & methods

2.1.   Samples

Milk samples from naturally occurring clinical cases of mastitis in individual udder quarters were obtained from the Vale

Veterinary Laboratory, Devon, UK with initial bacteriological culture performed in their laboratory to determine the bacterial cause of bovine mastitis. Results were confirmed by the Bacteriology Laboratory of the Veterinary Diagnostic Services of the School of Veterinary Medicine, University of Glasgow. Samples were plated on to 5% sheep blood agar and on to MacConkey agar for identification of *E. coli* while the presence of Gram positive bacteria including *S. aureus* was identified by API Staph and API 20 Strep tests kits (Biomerieux UK Ltd, Basingstoke, UK). Milk samples from cases of mastitis caused by either *E. coli* (n = 15) or *S. aureus* (n = 12) infection were selected for further analysis. Control milk samples (n = 10) from healthy cows from the School of Veterinary Medicine, University of Glasgow farm were used in this study as comparison to those clinical mastitic cases. The healthy milk samples were confirmed by somatic cell counts being < 100,000 cells/ml.

2.2. Sample preparation

Milk samples were defrosted at room temperature prior to centrifugation at 16,000 ×g for 10 min at 4 °C to remove the milk fat. An aliquot of 700 μl of centrifuged milk was added to 700 μl of 2 M urea, 100 mM NaCl, 10 mM NH_4OH containing 0.02% SDS, as described by Albalat et al. [21]. To remove higher molecular mass proteins the sample was ultrafiltered using Centrisart, Sartorius (Geottingen, Germany) ultracentrifugation filter devices (20 kDa MW) at 2000 g until 1.1 ml of filtrate was obtained. This filtrate was then applied onto a PD-10 desalting column, (GE Healthcare Biosciences, Uppsala, Sweden) equilibrated in 0.01% NH_4OH in HPLC-grade H_2O to decrease matrix effects by removing urea, electrolytes, salts, and to enrich polypeptides present. Finally, all samples were lyophilized and stored at 4 °C.

2.3. Protein estimation

Protein concentration in milk extracted samples was checked using the BCA assay Uptima from Interchim (Montluçon, France). A known amount (25 μl) of each standard and milk extracted samples was pipetted into microplate wells in duplicate. The BCA assay reagent was then added (200 μl) and mixed. The sample mixtures were incubated at 37 °C for 30 min. Prior to reading the results at the optical absorbance of 562 nm, the microplate was allowed to cool to room temperature. The standard curve of protein concentration, bovine serum albumin (Sigma Aldrich, Dorset, UK) ranging from 2 up to 500 μg/ml was plotted and the protein concentration in the sample was interpolated. Freeze-dried aliquots were re-suspended in appropriate volume of Milli Q water to reach a concentration of 2 μg/μl before running in the CE.

2.4. CE–MS analysis

Samples were re-suspended in HPLC-grade H_2O shortly before CE–MS analyses, as described [21]. CE–MS analysis was performed using a P/ACE MDQ capillary electrophoresis system (Beckman Coulter, Fullerton, USA) using a 90 cm 360 μm OD, 50 μm ID capillary with a tapered tip (New Objective, Woburn, USA) on-line coupled to a micrOTOF MS (Bruker Daltonic, Bremen, Germany) as described previously [22]. The ESI sprayer

(Agilent Technologies, Palo Alto, CA, USA) was grounded, and the ion spray interface potential was set between −4 and −4.5 kV. Data acquisition and MS acquisition methods were automatically controlled by the CE via contact-close-relays. Spectra were accumulated every 3 s, over a range of m/z 350 to 3000. Accuracy, precision, selectivity, sensitivity, reproducibility, and stability of the CE–MS measurements were demonstrated elsewhere [16].

2.5. LC–MS/MS analysis

The milk extracts were also analysed on a Dionex Ultimate 3000 RSLS nano-flow system (Dionex, Camberly UK). The samples (5 μl) were loaded onto a Dionex 100 μm × 2 cm 5 μm C18 nano-trap column at a flowrate of 5 μl/min by an Ultimate 3000 RS autosampler (Dionex, Camberley UK). The composition of the loading solution was 0.1% formic acid and acetonitrile (98:2). Once loaded onto the trap column the sample was then washed off into an Acclaim PepMap C18 nano-column 75 μm × 15 cm, 2 μm 100 Å at a flowrate of 0.3 μl/min. The trap and nano-flow column were maintained at 35 °C in a column oven in the Ultimate 3000 RSLC. The samples were eluted with a gradient of solvent A: 0.1% formic acid and acetonitrile (98:2) versus solvent B: 0.1% formic acid and acetonitrile (20:80) starting at 5% B rising to 50% B over 100 min. The column was washed using 90% B before being equilibrated prior to the next sample being loaded.

The eluant from the column was directed to a Proxeon nano-spray ESI source (Thermo Fisher Hemel UK) operating in positive ion mode then into an Orbitrap Velos FTMS. The ionisation voltage was 2.5 kV and the capillary temperature was 200 °C. The mass spectrometer was operated in MS–MS mode scanning from 380 to 2000 amu. The top 20 multiply charged ions were selected from each full scan for MS–MS analysis, the fragmentation method was CID at 35% collision energy. The ions were selected for MS^2 using a data dependant method with a repeat count of 1 and repeat and exclusion time of 15 s. Precursor ions with a charge state of 1 were rejected. The resolution of ions in MS^1 was 60,000 and 7500 for HCD MS^2.

2.6. CE–MS data processing

Mass spectra were processed using MosaiquesVisu software (Mosaiques Diagnostics, Hannover, Germany), including peak picking, deconvolution and deisotoping [23]. The software automatically examines all mass spectra from a CE–MS analysis for signals above the threshold (SNR ^4). Only signals that are present in 3 consecutive spectra are accepted. Next, the isotopic distribution is assessed and charge is assigned based on the isotopic distribution, using a matched filtering algorithm. This operation results in a list wherein all signals that could be interpreted are defined by mass/charge, charge, migration time, and signal intensity (ion counts). This list is transformed into a dataset containing only mass, migration time, and signal intensity, signals that represent the same compound but with a different charge state are combined. The final result is a list of the compounds present in the sample with their identity defined by mass, migration time and their relative abundance defined by ion counts.

CE migration time and ion signal intensity were normalized to allow compilation and comparison of samples. Reference signals of over 380 'milk housekeeping polypeptides' were used for CE-time calibration by local regression. The same peptides were used for ion signal intensity normalization by a applying a global linear regression. The resulting peak list contained the molecular mass (Da) and normalized CE migration time (min) for each feature. Normalized signal intensity can be used as a measure of relative abundance. Data sets were accepted only if they passed a strict quality control criteria: A minimum of 950 chromatographic features (mean number of features minus one standard deviation) must be detected with a minimal MS resolution of 8000 (required resolution to resolve ion signals with $z = 6$) in a minimal migration time interval (the time window, in which separated signals can be detected) of 10 min. After calibration, the mean deviation of migration time (compared to reference standards) must be below 0.35 min.

Control and disease-specific polypeptide patterns were generated using support vector machine (SVM)-based MosaCluster software Mosaiques Diagnostics (Hannover, Germany).

2.7.   LC–MS/MS—data processing

Raw spectral data from LC–MS/MS analysis of the samples were uploaded to Thermo Proteome Discoverer 1.2 Thermo Scientific (Hemel Hempstead, UK). Peak picking was performed under default settings for FTMS analysis such that only peptides with signal to noise ratio higher than 1.5 and belonging to precursor peptides between 700–8000 Da were considered. Peptide and protein identification was performed with SEQUEST algorithm. An in house compiled database containing proteins from the latest version of the UniProt SwissProt database was compiled to include only *Bos taurus*, *E. coli* and *S. aureus* entries. This concatenated database was selected to facilitate peptide identification and reduce the probability for false positive identifications upon making searches against 3 differently sized FASTA databases. No enzyme cleavage was selected and oxidation of methionine and proline was chosen as variable modifications. Precursor tolerance was set at 20 ppm and 0.1 Da for MS/MS fragment ions. Resulting peptides and protein hits were further screened by excluding peptides with an error tolerance higher than 10 ppm and by accepting only those hits listed as high confidence by Proteome Discoverer software. Theoretical migration times in CE–MS for any resulting peptides were calculated so that sequences obtained with LC-MS/MS could be subsequently assigned to a position in the CE–MS analysis.

3.   Results

3.1.   Polypeptide fingerprinting for control and infected groups

In each sample analysed we were able to detect >1000 different peptides based on migration time and mass. The molecular weight range of these peptides ranged from 0.8 to 20 kDa. Data from individual CE–MS analysis was compiled according to their group in order to develop biomarker patterns for the different conditions studied.

3.1.1.   Healthy versus mastitis samples

Initially in this work we investigated whether differences existed in the milk proteome of healthy cows (n = 10) when compared to mastitic cows, regardless of their bacterial cause (n = 27). The polypeptide fingerprints of these two groups were very different as shown in Fig. 1, with many of the multiply charged high molecular weight peptides being absent from the infected group. A total of 508 peptides were significantly different after correcting for multiple testing (BH < 0.05).
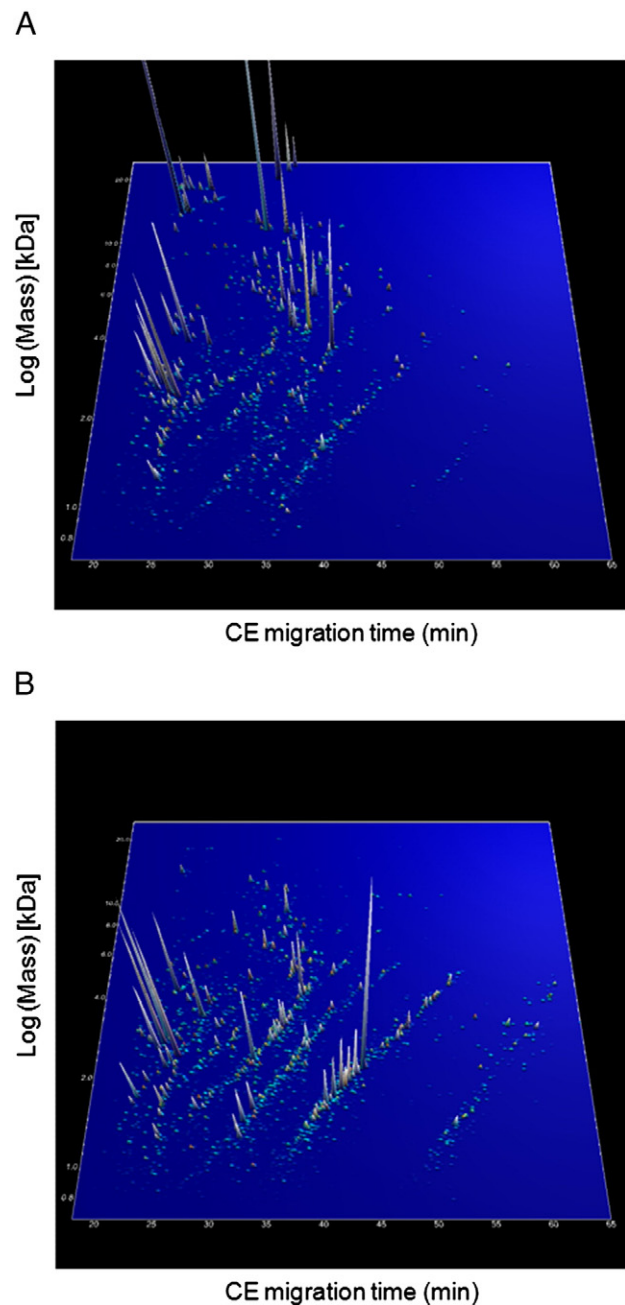

Fig. 1 – Compiled CE–MS peptide fingerprints of the control milk samples (n = 10) (A) and all mastitic milk samples (n = 27) (B). On the X axis the CE migration time (min) is plotted against the molecular mass (kDa) on the Y axis on a logarithmic scale. The Z axis represents the mean signal intensity.

3.1.2. Model generation of the polypeptide fingerprints

Based on the data described above, multiple testing of statistical analysis was done to discriminate peaks between control and infected groups. In the first model, using a training set of randomly selected controls (n = 5) and infected (n = 14) samples we selected discriminatory peptides with a BH < 0.05 and with the area under curve (AUC) set at 1. Following these very stringent conditions a group of 154 of the 508 peaks was determined to be a potential biomarker panel that can be used to discriminate between control and infected groups. Upon using leave-one-out cross-validation of the training set for verification an accuracy of 100% was observed. This model was applied to a test set of controls (n = 5) and infected (n = 13) with a sensitivity and specificity of 100%. Therefore, peptides from this model could be used to discriminate between non-infected and infected group with an accuracy of 100% (Fig. 2).

3.1.3. E. coli infected versus S. aureus infected samples

The next step was to investigate the differences in milk proteome of infected samples caused by the two different bacterial pathogens (E. coli versus S. aureus). The polypeptide profiles of these 2 different groups did not differ as markedly as is demonstrated in Fig. 3. A total of 47 peptides were significantly different between both groups after correcting for multiple testing (BH < 0.05). In this case the biomarker panel was made up of all 47 discriminatory peptides.

3.1.4. Model generation of the polypeptide fingerprints

Based on the polypeptide fingerprints described above, multiple testing of statistical analysis was done to discriminate peaks between two infected groups caused by E. coli and S. aureus. A model was generated using a training set of un-selected samples from E. coli (n = 11) and S. aureus (n = 8) and a group of 47 peptides (significantly different peptides after correcting for multiple testing) was determined to be a biomarker panel that could discriminate between causative bacteria. Upon cross-validation of the training set an accuracy of 100% was obtained. This model was then tested in a test set
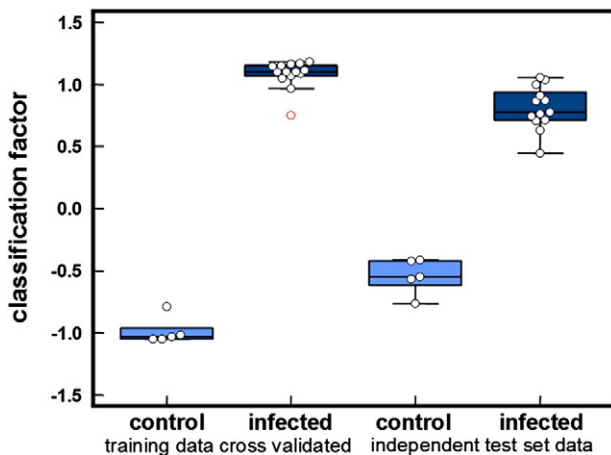

Fig. 2 – Box-and-whisker plots of the classification factor obtained for the control (n = 5) vs. infected (n = 14) of the training set and the test set in a blinded assessment of milk samples from control (n = 5) and infected (n = 13) cows.
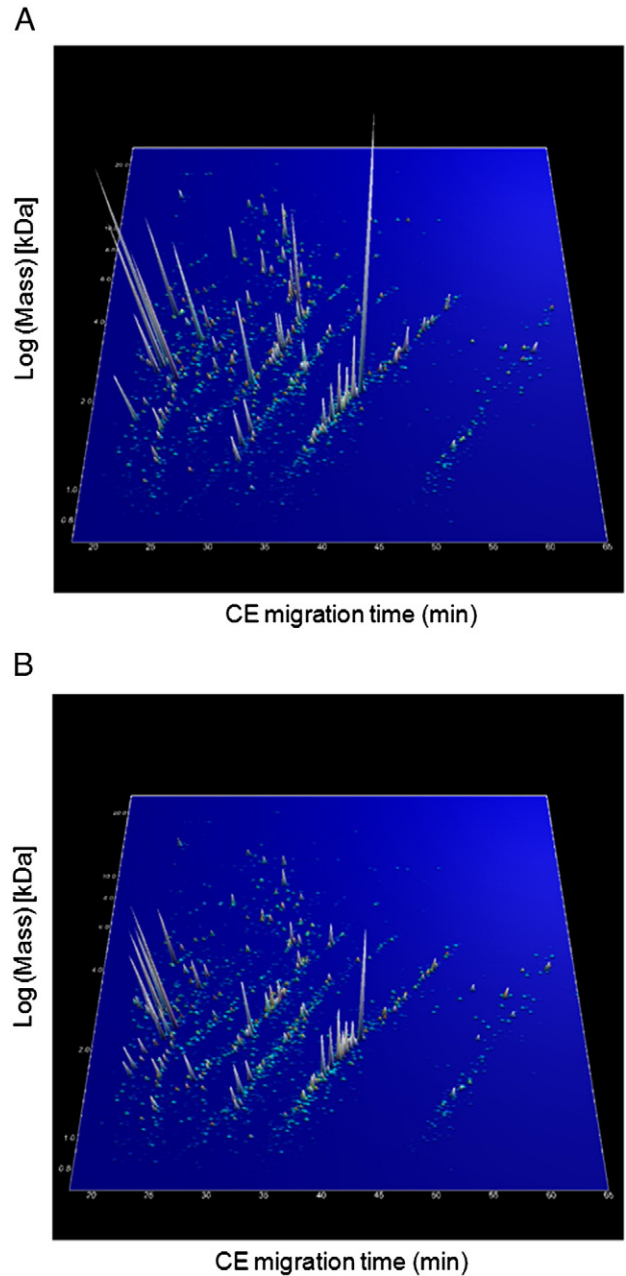

Fig. 3 – Compiled CE–MS peptide fingerprints of the milk samples from quarters with mastitis caused by E. coli (n = 15) (A) and S. aureus (n = 12) (B) infections. On the X axis the CE migration time (min) is plotted against the molecular mass (kDa) on the Y axis on a logarithmic scale. The Z axis represents the mean signal intensity.

of E. coli (n = 4) and S. aureus (n = 4). This model allowed a correct classification of 4/4 E. coli and 3/4 S. aureus thus giving a sensitivity on the limited test set of 75% while the specificity was 100% (Fig. 4).

3.2. Peptide sequencing

In order to identify the discriminatory peptides of the models by their sequences, samples were analysed using LC–MS/MS. The
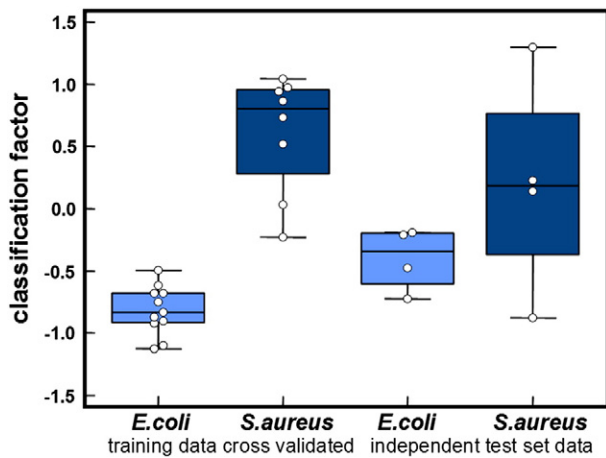
Fig. 4 – Box-and-whisker plots of the classification factor obtained for the *E. coli* (n = 11) vs. *S. aureus* (n = 8) of the training set and the test set in a blinded assessment of infected milk samples from *E. coli* (n = 4) and *S. aureus* (n = 4) samples.

data was then compared with that of the UniProt SwissProt database which was compiled to include only *Bos taurus*, *E. coli* and *S. aureus* entries. All identified peptides obtained with high confidence from the LC–MS/MS analysis, that matched the mass and migration time of a peptide in the CE–MS analysis used in the models, are listed in Tables 1 and 2. In the model classifying healthy versus infected samples 33 peptides were identified out of 154 peptides (Table 1). On the other hand, from the model used to classify samples from *E. coli* versus *S. aureus* 15 peptides were sequenced out of the 47 that formed the model (Table 2). The frequency of occurrence of each peptide is listed in Tables 1 and 2 as well as their mean amplitude.

3.3. Protein identification

As described above, the protein identity was derived from matching the amino acid sequences against a protein database. The identified and named proteins are presented as in Tables 1 and 2. Combining both sets, peptides originating from twelve proteins were identified with the majority of the originating proteins being known milk protein such as α_{S1}-casein and β-casein.

4.   Discussion

The chosen peptidomic method of capillary electrophoresis (CE) coupled with the electrospray mass spectrometry is known for its rapid (approximately 45 min per sample), sensitive and automated approach [24]. The principle of separation in a CE system ensures that all peptides present in the sample will migrate through the capillary and pass into the detector in a consistent solvent suitable for detection. This analytical platform was used to detect quantitative differences in the peptidome of milk between samples of milk from healthy cows (control) and milk from infected udder quarters of cows with mastitis caused by two different bacterial pathogens (*E. coli* and *S. aureus*).

All milk samples were analysed by CE–MS and only poly-peptides present at a frequency >80% in at least one group (control or cases) were chosen for analysis. The milk poly-peptide patterns of infected cows were markedly different from control cows. The milk polypeptide patterns when comparing the causative bacteria (*E. coli* and *S. aureus*) were more similar but still had identifiable quantifiable differences.

There are only a very small number of studies in which peptide analysis has been performed on bovine milk samples during mastitis [15,25]. Both studies employed the same method, RP-HPLC, in separating the peptides, while further protein characterization was performed by MALDI-TOF/MS/MS. These studies focused on the result of proteolysis caused by different types of proteolytic enzymes. Wedholm et al. [25] successfully identified 10 peptides from milk with high SCC (16 000 000 cells/ml). Larsen et al. [15] identified approximately 20 different peptides in milk samples from mammary glands infused with LTA from *S. aureus*, with peptides resulting from the action of endogenous proteases in response to the LTA. The objective of these studies was to investigate milk protease activity during mastitis. In contrast, the objective of the present study was to determine how the peptide profile of milk samples is altered in bovine mastitis. The CE–MS methodology used in this study allows quantitative analysis of all the peptides present in the samples. It has a number of advantages in peptide analysis over the more popular methodology of LC separation followed by MS or MS/MS identification [26].

A secondary objective was to discover and characterize the peptides that made up the biomarker in bovine mastitis. However, due to the small sample loading capacity in CE it is not possible to obtain MS/MS data to allow sequencing. Sequencing was carried out using LC–MS/MS. LC separation of peptides misses many of the most polar peptides and also the larger more hydrophilic molecules and therefore many of the peptides used to generate the biomarker panels have no sequence data. However, we did see some overlap with the 14 peptide sequences described by Larsen et al. [15]. In Table 1 we have no exact matches though peptide IDs 2367, 2719, 3223 and 3879 are smaller fragments of one of the peptides listed by Larsen et al. [15]. In Table 2 there is one exact match with another peptide described by Larsen, peptide ID 10468 and a smaller fragment of this peptide ID 9569. As the majority of peptides in the biomarkers are only seen in CE–MS and our sequencing is by LC–MS/MS it is reassuring that there is some overlap with the LC fractions identified by MALDI/MS/MS analysis of Larsen et al. [15]. From the results it is clear that in clinical mastitis the milk proteome, or more precisely the milk peptidome as only peptides of <20 kDa were examined in this study, increases in complexity with an increase in the number and intensity of low MW peptides. The high sensitivity and specificity of the differentiation of milk from healthy animals compared to that from udder quarters suffering from mastitis were an encouraging result in biomarker discovery for this economically important disease. Furthermore the potential for discrimination between the causative bacteria by peptide analysis is an exciting development. Currently, such discrimination would require time consuming bacterial identification. If a rapid method could be developed, following validation of discrimination in future studies, there could be

Table 1 – Protein identities from peptides used in the model to differentiate healthy (control) versus infected (case) samples. Model was composed of 154 peptides from which 33 were identified.

Peptide ID	Peptide mass (Da)	CE time (min)	Sequence peptide	Protein identity	Freq control	Freq case	Mean amplitude control	Mean amplitude case
479	932.53	35.34	LIVTQTMK	Major allergen β-lactoglobulin	1.00	0.86	2.79	1.41
922	1019.58	35.96	EPVLGPVRGP	β-casein	1.00	0.29	2.29	0.25
1299	1093.66	33.86	PVRGPFPIIV	β-casein	1.00	0.29	1.62	0.10
2367	1275.67	32.18	PFPEVFGKEKV	α_{S1}-casein	0.60	1.00	0.51	3.50
2426	1286.70	30.80	QVWEESLKRL	Lactoperoxidase	1.00	0.07	1.47	0.03
2644	1332.65	37.83	EMPFPKYPVEP	β-casein	–	1.00	–	1.98
2719	1346.72	31.92	VAPFPEVFGKEK	α_{S1}-casein	0.80	1.00	1.57	3.64
2757	1652.76	21.06	SKVKEAMAPKHK	β-casein	0.20	1.00	0.23	2.28
3058	1406.71	38.68	LWYNMLVAEPR	Signal transducer and activator of transcription 1	1.00	–	1.24	–
3223	1445.79	32.44	VAPFPEVFGKEKV	α_{S1}-casein	–	1.00	–	3.87
3864	1588.92	36.06	EPVLGPVRGPFPIIV	β-casein	1.00	0.14	2.48	0.21
3879	1592.86	33.25	FVAPFPEVFGKEKV	α_{S1}-casein	–	1.00	–	3.83
3882	1593.70	39.49	RSNVQSPDATEEDF	Osteopontin	1.00	–	2.46	–
3913	1597.81	27.98	HKEMPFPKYPVEP	β-casein	–	1.00	–	3.01
4226	1667.90	40.05	LYQEPVLGPVRGPFP	β-casein	–	1.00	–	3.865
4394	1713.80	40.49	IPNPIGSENSEKTTMP	α_{S1}-casein	0.20	1.00	0.07	2.64
4639	1780.98	40.76	LYQEPVLGPVRGPFPI	β-casein	0.40	1.00	0.46	3.95
4710	1799.03	29.26	YLGYLEQLLRLKKY	α_{S1}-casein	–	1.00	–	3.00
5067	1894.08	41.47	LYQEPVLGPVRGPFPII	β-casein	0.20	1.00	0.28	3.67
5385	1978.95	34.68	RAAPGPASAPSSGYREFVQ	Interleukin 4-receptor	–	1.00	–	1.77
5445	1997.06	26.69	GSKASADESLALGKPGKEPR	Fibroblast growth factor-binding protein	1	0.21	2.28	0.35
5482	2007.15	42.04	LLYQEPVLGPVRGPFPII	β-casein	0.20	1.00	0.28	2.95
5594	2041.14	41.99	FLLYQEPVLGPVRGPFPI	β-casein	–	1.00	–	2.93
5904	2125.03	42.75	LQSQASRSTISSSFGNEETP	Rho guanine nucleotide exchange factor	0.20	1.00	0.42	3.84
6017	2153.17	24.36	RGSKASADESLALGKPGKEPR	Fibroblast growth factor-binding protein	1.00	0.64	2.55	0.71
6158	2190.06	32.31	HKEMPFPKYPVEPFTESQ	β-casein	–	1.00	–	2.28
6240	2215.05	43.51	SDIPNPIGSENSEKTTMPLW	α_{S1}-casein	1.00	1.00	1.85	4.28
6300	2231.06	43.73	SDIPNPIGSENSEKTTmPLW	α_{S1}-casein	–	1.00	–	2.85
6465	2277.28	27.64	SSRQPQSQNPKLPLSILKEK	Glycosylation-dependent cell adhesion molecule	1.00	0.71	3.27	0.82
6961	2406.17	33.25	HKEMPFPKYPVEPFTESQSL	β-casein	–	1.00	–	2.44
7095	2440.39	24.84	SRQPQSQNPKLPLSILKEKHL	Glycosylation-dependent cell adhesion molecule	1.00	0.14	2.26	0.14
7454	2527.42	25.01	SSRQPQSQNPKLPLSILKEKHL	Glycosylation-dependent cell adhesion molecule	1.00	0.86	4.22	1.56
11162	3464.84	35.54	HLPLPLLQSWMHQPHQPLPPTVMFPPQSVL	β-casein	–	1.00	–	2.35

major benefits in targeting treatment to the correct pathogen. Whether the differences in peptidome pattern are due directly to bacterial action in breaking down larger milk proteins or are due to differences in the host response to infection is not clear but should be the objective of future investigation.

A total of 48 polypeptides were sequenced by LC–MS/MS (Tables 1 & 2). These polypeptides were identified based on their protein accession information and amino acid sequences and were found to be fragments of larger and mostly known milk proteins. The UniPortKB database was used to search for each of the protein identities. The result of the database search using the peptide sequence produced a list of proteins including α_{S1}-casein, β-casein, β-lactoglobin, lactoperoxidase, osteopontin, interleukin 4-receptor, fibroblast growth factor binding protein, glycosylation-dependent cell adhesion molecule 1 (GlyCAM-1), rho guanine nucleotide exchange factor,

signal transducer and activator of transcription 1 and tubulin α-1C chain. One peptide produced two possible sequences which matched both mass and migration time from the CE–MS data. The proteins identified from these two potential sequences were β-casein and ANKRD9 protein. As we are interested in discriminatory polypeptides, the discussion will focus on the polypeptides and related proteins that were used to discriminate between control, infected and two different bacterial pathogen groups.

Statistically, peptides derived from β-lactoglobin, α S1-casein, β-casein, lactoperoxidase, osteopontin, interleukin 4 receptor, FGFBP and glyCAM 1 were the peptides that can be used to discriminate the control from the infected groups.

During mastitis, the endogenous proteolytic activity in milk increases substantially [15,27,28] and it is known that plasmin can hydrolyse β-casein, α_{S2}-casein and α_{S1}-casein [29] which

Table 2 – Protein identities from peptides used in the model to differentiate *E. coli* versus *S. aureus*. Model was composed of 47 peptides from which 15 were identified.

Peptide ID	Peptide mass (Da)	CE time (min)	Sequence peptide	Protein identity	Freq E. coli	Freq S. Aureus	Mean amplitude E. coli	Mean amplitude S. aureus
670	975.53	36.14	FPKYPVEP	β-casein	1.00	0.67	1.62	0.64
3264	1454.77	27.56	SETGAGKHVPRAVF	Tubulin α-1C chain	0.13	1.00	0.07	2.07
3488	1495.77	39.17	RELEELNVPGEIV	β-casein	1.00	0.20	1.36	0.43
3869	1589.84	33.56	FLQPEVMGVSKVKE	β-casein	0.50	1.00	0.56	2.17
4247	1673.86	39.47	VPYPQRDMPIQAFL	β-casein	1.00	0.17	2.03	0.25
4582	1763	22.30	RPKHPIKHQGLPQEV	α_{S1}-casein	1.00	0.83	3.98	2.98
5593	2041.06	26.54	APKHKEMPFPKYPVEPF	β-casein	0.25	0.83	0.47	1.43
7524	2546.30	26.50	EERLHSMKEGIHAQQKEPMIGV	α_{S1}-casein	1.00	0.17	2.57	0.31
8016	2656.44	26.04	KHPIKHQGLPQEVLNENLLRFF	α_{S1}-casein	1.00	0.33	2.47	0.41
8093	2676.48	34.40	SQSKVLPVPQKAVPYPQRDMPIQA	β-casein	0.38	1.00	0.31	2.91
8639	2801.40	38.80	RELEELNVPGEIVESLSSSEESITR	β-casein	1.00	0.50	2.23	0.93
9569	3005.65	36.49	LSLSQSKVLPVPQKAVPYPQRDMPIQA	β-casein	1.00	0.67	2.36	1.00
9725	3044.48	31.21	KIEKFQSEEQQQTEDELQDKIHPFA	β-casein	0.88	0.33	2.33	0.45
10468	3249.83	35.90	LSLSQSKVLPVPQKAVPYPQRDMPIQAFL	β-casein	1.00	0.50	3.96	0.78
10479	3252.67	35.15	HLPLPLLQSWMHQPHQPLPPTVMFPPQS	β-casein	0.88	0.50	2.66	0.86

explains the presence of peptides from α_{S1}-casein and β-casein in milk from the infected group. The results presented here, from investigation of naturally occurring mastitis, are in agreement with investigations of experimental induction of mastitis with LTA [15] or lipopolysaccharide [28] in which plasmin activity in milk increased and casein derived peptides were identified by reverse phase HPLC [15] or 2 dimension electrophoresis and mass spectroscopy [28] during mastitis. It is likely that plasmin and other endogenous peptides are responsible in part at least for the production of the peptides identified by this investigation. As α_{S1}-casein and β-casein are major milk proteins [30,31] they constitute a principle substrate for milk proteolysis resulting from the action of endogenous proteinases such as plasmin [32]. Although proteolytic activity in milk is known to increase as a result of mastitis, other factors such as the stage of lactation and lactation number can also contribute to increase in this proteolytic activity [33].

The presence of GlyCAM 1 protein in bovine milk has been reported previously [11,34]. Smolenski et al. [11] have identified it as one of the minor milk protein that is present in all peak lactation, colostrums, mastitic whey and the milk fat globule membrane (MFGM) whilst Boehmer et al. [34] considered it to be a host defence protein. It was surprising to observe this minor milk protein which acts in the host defence mechanism to appear in the control group in higher frequency compared with the infected group.

The fibroblast growth factor-binding protein (FGFBP) is a secreted carrier protein that releases fibroblast growth factors (FGFs) from the extracellular matrix storage and thus enhancing the FGF activity [35]. Kawakami et al. [36] concluded that FGFBP may be tightly associated with lactoferrin in bovine milk. Plath et al. [37] have successfully examined the expression and localization of FGFBP in the bovine mammary gland and it may be important in the local regulation of the bovine mammary gland.

Statistically, peptides from α_{S1}-casein, β-casein and tubulin α-1C chain protein were the peptides that can be used to discriminate between the bacterial cause of the mastitis.

β-casein derived peptides were assigned to discriminate *S. aureus* from *E. coli* infected group. As discussed above, the presence of mastitis can increase the proteolytic activity in bovine milk and it was not surprising that these polypeptides were the result from β-casein hydrolysis, but it is not clear how a number of β-casein derived peptides can contribute to differential identification of the bacterial cause while other β-casein peptides are included in the group to differentiate milk from mastitic from healthy udders. It was also not clear if β-casein peptides alone can be applied to discriminate *S. aureus* from *E. coli* infected milk samples.

Although several proteins were successfully identified using the protein database based on their sequences, those proteins cannot be used per se as potential markers for each group. This study only sequenced peptides resulting from specific fragments of proteins. The peptide identity provides an insight into which milk proteins were subjected to proteolysis or played a role in the disease mechanism. It is apparent that specific peptides resulting from particular proteases of the host or bacteria were of varying importance in contributing to the discriminatory algorithms. However, no ultimate conclusion can be made based on the protein identity that has been identified in each of the group as the group sizes are relatively small. Further study of more extensive numbers of samples, stage of disease (clinical and subclinical) and causative bacteria is warranted to determine if these promising results are indicative of a viable approach to improve the diagnosis of bovine mastitis. Eventual use of the panel of peptide found here in differential diagnosis of mastitis could have a significant impact on the health of dairy cows and in control of this important disease [3] allowing more efficient targeting of antibiotic use. It would also be interesting to investigate the time-course of infection and determine the earliest time that can be applied to diagnose an infected cow based on the proposed polypeptide biomarker of bovine mastitis. More information on peptide sequencing will also lead to a better understanding of the pathophysiology of bovine mastitis.

Acknowledgements

REFERENCES

[1] Radostits OM, Blood DC, Gay CC, Hinchkliff KW. Veterinary medicine. 9th ed.London: ELBS-Bailliere Tindal; 2000.

[2] Contreras GA, Rodriguez JM. Mastitis: comparative etiology and epidemiology. J Mammary Gland Biol 2011;16:339–56.

[3] De Vliegher S, Fox L, Piepers S, McDougall S, Barkema H. Invited review: mastitis in dairy heifers: nature of the disease, potential impact, prevention, and control. J Dairy Sci 2012;95:1025–40.

[4] Schukken YH, Guenther J, Fitzpatrick J, Fontaine M, Goetze L, Holst O, et al. Host-response patterns of intramammary infections in dairy cows. Vet Immunol Immunopathol 2011;144:270–89.

[5] Halasa T, Huijps K, Osteras O, Hogeveen H. Economic effects of bovine mastitis and mastitis management: a review. Vet Q 2007;29:18–31.

[6] Miller GY, Bartlett PC, Lance SE, Anderson J, Heider LE. Costs of clinical mastitis and mastitis prevention in dairy herds. J Am Vet Med Assoc 1993;202:1230–6.

[7] Biggs A. Mastitis in cattle. Marlborough, Wiltshire, UK: Crowood Press; 2009.

[8] Seegers H, Fourichon C, Beaudeau F. Production effects related to mastitis and mastitis economics in dairy cattle herds. Vet Res 2003;34:475–91.

[9] Wellnitz O, Bruckmaier RM. The innate immune response of the bovine mammary gland to bacterial infection. Vet J 2012;192:148–52.

[10] Hogarth CJ, Fitzpatrick JL, Nolan A, Young FJ, Pitt A, Eckersall PD. Differential protein composition of bovine whey: a comparison of whey from healthy animals and from those with clinical mastitis. Proteomics 2004;4:2094–100.

[11] Smolenski G, Haines S, Kwan FYS, Bond J, Farr V, Davis SR, et al. Characterisation of host defence proteins in milk using a proteomic approach. J Proteome Res 2007;6:207–15.

[12] Boehmer JL, Bannerman DD, Shefcheck K, Ward JL. Proteomic analysis of differentially expressed proteins in bovine milk during experimentally induced Escherichia coli mastitis. J Dairy Sci 2008;91:4206–18.

[13] Danielsen M, Codrea MC, Ingvartsen KL, Friggens NC, Bendixen E, Rontved CM. Quantitative milk proteomics—host responses to lipopolysaccharide-mediated inflammation of bovine mammary gland. Proteomics 2010;10:2240–9.

[14] Ibeagha-Awemu EM, Ibeagha AE, Messier S, Zhao X. Proteomics, genomics, and pathway analyses of Escherichia coli and Staphylococcus aureus infected milk whey reveal molecular pathways and networks involved in mastitis. J Proteome Res 2010;9:4604–19.

[15] Larsen T, Hinz K, Jorgensen AL, Moller HS, Wellnitz O, Bruckmaier RM, et al. Proteomic and peptidomic study of proteolysis in quarter milk after infusion with lipoteichoic acid from Staphylococcus aureus. J Dairy Sci 2010;93:5613–26.

[16] Good DM, Zuerbig P, Argiles A, Bauer HW, Behrens G, Coon JJ, et al. Naturally occurring human urinary peptides for use in diagnosis of chronic kidney disease. Mol Cell Proteomics 2010;9(11):2424–37.

[17] Metzger J, Kirsch T, Schiffer E, Ulger P, Mentes E, Brand K, et al. Urinary excretion of twenty peptides forms an early and accurate diagnostic pattern of acute kidney injury. Kidney Int 2010;78(12):1252–62.

[18] Delles C, Schiffer E, von zur Muhlen C, Peter K, Rossing P, Parving HH, et al. Urinary proteomic diagnosis of coronary artery disease: identification and clinical validation in 623 individuals. J Hypertens 2010;28(11):2316–22.

[19] Maahs DM, Siwy J, Argiles A, Cerna M, Delles C, Dominiczak AF, et al. Urinary collagen fragments are significantly altered in diabetes: a link to pathophysiology. PLoS One 2010;5(9).

[20] Siwy J, Mullen W, Golovko I, Franke J, Zuerbig P. Human urinary peptide database for multiple disease biomarker discovery. Proteomics Clin Appl 2011;5:367–74.

[21] Albalat A, Franke J, Gonzalez J, Mischak H, Zurbig P. Urinary proteomics based on capillary electrophoresis coupled to mass spectrometry in kidney disease. Methods Mol Biol 2013;919:203–13.

[22] Zurbig P, Schiffer E, Mischak H. Capillary electrophoresis coupled to mass spectrometry for proteomic profiling of human urine and biomarker discovery. Methods Mol Biol 2009;564:105–21.

[23] Neuhoff NV, Kaiser T, Wittke S, Krebs R, Pitt A, Burchard A, et al. Mass spectrometry for the detection of differentially expressed proteins: a comparison of surface-enhanced laser desorption/ionization and capillary electrophoresis/mass spectrometry. Rapid Commun Mass Spectrom 2004;18:149–56.

[24] Kolch W, Neususs C, Peizing M, Mischak H. Capillary electrophoresis–mass spectrometry as a powerful tool in clinical diagnosis and biomarker discovery. Mass Spectrom Rev 2005;24:959–77.

[25] Wedholm A, Moller HS, Stensballe A, Lindmark-Mansson H, Karlsson AH, Andersson R, et al. Effect of minor milk proteins in chymosin separated whey and casein fractions on cheese yield as determined by proteomics and multivariate data analysis. J Dairy Sci 2008;91:3787–97.

[26] Mullen W, Albalat A, Gonzalez J, Zerefos P, Siwy J, Franke J, et al. Performance of different separation methods interfaced in the same MS-reflection TOF detector: a comparison of performance between CE versus HPLC for biomarker analysis. Electrophoresis 2012;33:567–74.

[27] Verdi RJ, Barbano DM, Dellavalle ME, Senyk GF. Variability in true protein, casein, nonprotein nitrogen, and proteolysis in high and low somatic-cell milks. J Dairy Sci 1987;70:230–42.

[28] Hinz K, Larsen LB, Wellnitz O, Bruckmaier RM, Kelly AL. Proteolytic and proteomic changes in milk at quarter level following infusion with Escherichia coli lipopolysaccharide. J Dairy Sci 2012;95:1655–66.

[29] Larsen LB, McSweeney PLH, Hayes MG, Andersen JB, Ingvartsen KL, Kelly AL. Variation in activity and heterogeneity of bovine milk proteases with stage of lactation and somatic cell count. Int Dairy J 2006;16:1–8.

[30] O'Donnell R, Holland JW, Deeth HC, Alewood P. Milk proteomics. Int Dairy J 2004;14:1013–23.

[31] Manso MA, Leonil J, Jan G, Gagnaire V. Application of proteomics to the characterisation of milk and dairy products. Int Dairy J 2005;15:845–55.

[32] Bastian ED, Brown RJ. Plasmin in milk and dairy products: an update. Int Dairy J 1996;6:435–57.

[33] Politis I, Lachance E, Block E, Turner JD. Plasmin and plasminogen in bovine-milk—a relationship with involution. J Dairy Sci 1989;72:900–6.

[34] Boehmer J, Ward J, Peters R, Shefcheck K, McFarland M, Bannerman D. Proteomic analysis of the temporal expression of bovine milk proteins during coliform mastitis and label-free relative quantification. J Dairy Sci 2010;93:593–603.

[35] Xie B, Tassi E, Swift MR, McDonnell K, Bowden ET, Wang SM, et al. Identification of the fibroblast growth factor (FGF)-interacting domain in a secreted FGF-binding protein by phage display. J Biol Chem 2006;281:1137–44.

[36] Kawakami A, Hirayama K, Kawakami F, Kawakami H, Fujihara M, Ohtsuki K. Purification and biochemical characterization of a fibroblast growth factor-binding protein (FGF-BP) from the lactoferrin fraction of bovine milk. BBA-Gen Subj 2006;1760:421–31.

[37] Plath A, Einspanier R, Gabler C, Peters F, Sinowatz F, Gospodarowicz D, et al. Expression and localization of members of the fibroblast growth factor family in the bovine mammary gland. J Dairy Sci 1998;81:2604–13.