International Dairy Journal 117 (2021) 104990

Contents lists available at ScienceDirect

International Dairy Journal

journal homepage: www.elsevier.com/locate/idairyj

Endogenous protein and peptide analysis with LC-MS/(MS): A feasibility study for authentication of raw-milk farmer's cheese

Ioana M. Barbu [a, *, 1], Robert-Jan A.N. Lamers [b, 1], Henk W. Gerritsen [a], Marco H. Blokland [a], Monique G.E.G. Bremer [a], Martin Alewijn [a]

[a] Wageningen Food Safety Research (WFSR), Wageningen University & Research, P.O. Box 230, 6700, AE, Wageningen, the Netherlands
[b] Abundnz B.V., Noordveenweg 3c, 2376, AP, Nieuwe Wetering, the Netherlands

ARTICLE INFO

Article history:
Received 27 July 2020
Received in revised form
17 December 2020
Accepted 4 January 2021
Available online 23 January 2021

ABSTRACT

When consumers buy Dutch 'Boerenkaas' (farmer's cheese), a cheese made from raw milk protected under the European traditional specialties guaranteed (TSG) label, they expect this product to be authentic. Because of the difference in production costs and risks, it is tempting to sell cheeses made from heat-treated milk as 'Boerenkaas'. It is therefore essential that there are methods that can verify the heat-treatment status of the milk in cheese. Here, for the first time a proof of principle method for analysis of endogenous cheese proteins/peptides with liquid chromatography-mass spectrometry (LC-MS) in combination with multivariate analysis to detect discriminatory protein/peptide biomarkers between 'Boerenkaas' and heat-treated milk cheese is described. A top-down MS method was developed for biomarker identification. The identified biomarkers were confirmed with the well-established bottom-up approach. Overall, there was good agreement between the outcome of the developed methods based on the known thermal treatment of the used sample set.

1. Introduction

'Boerenkaas', a Gouda-type raw milk cheese, is a product protected by the European traditional speciality guaranteed (TSG) label (European Commission, 2009). One of the prerequisites is that these cheeses should be made from milk, the temperature of which never exceeds 40 °C. For reasons described previously (Alewijn, Wehrens, & van Ruth, 2018), it is conceivable that cheeses being marketed as 'Boerenkaas' are in fact not, where (partial) heat-treatment of the milk before cheese-making is a practise that can easily be applied. Methodologies that can determine the heating status of the milk in the cheese, regardless of the exact cheese-making process and changes during ripening, are valuable to verify the authenticity of 'Boerenkaas' on the market.

The application of raw milk is important for the authenticity of many farmhouse kinds of cheese. Several heat treatment indicators and their detection methods in milk have been described, such as heat-sensitive enzymes native to milk (e.g., alkaline phosphatase, peroxidase and γ-glutamyl transpeptidase (Griffiths, 1986; McKellar, Emmons, & Farber, 1991), and specific indicators such as lactulose, furosine and hydroxymethylfurfural (Claeys, Van Loey, & Hendrickx, 2002), specific proteins such as whey proteins (Boitz, Fiechter, Seifried, & Mayer, 2015; Lin, Sun, Cao, Cao, & Jiang, 2010; Recio, Amigo, & Lopez-Fandino, 1997), denatured proteins and Maillard reaction end products (Birlouez-Aragon, Sabat, & Gouti, 2002), specific peptides in milk (Ebner, Baum, & Pischetsrieder, 2016), and other forms of protein modification (Johnson, Philo, Watson, & Mills, 2011). From the category of methods that use heat-sensitive milk enzymes (Griffiths, 1986) as indicators for heat treatment, alkaline phosphatase is well-known and standardised (ISO 11816). Some of the markers used for the detection of raw and heated milk are affected in the process of cheese-making, and thus not all methods above are suitable to detect the heat treatment status of cheese milk in the final product (cheese).

Alternative methods, such as those that detect volatile organic compounds in cheese using, for example, e-nose or proton transfer reaction-mass spectroscopy (PTR-MS), have been developed (Alewijn et al., 2018; Bergamaschi et al., 2016; Gasperi et al., 2001) and can be used to discriminate cheese made from heated or raw milk.

* Corresponding author. Tel.: +31 317 487608.
E-mail address: ioana.barbu@wur.nl (I.M. Barbu).
[1] These authors contributed equally.

https://doi.org/10.1016/j.idairyj.2021.104990
0958-6946/

Thermal treatment of milk induces various physicochemical modifications in milk molecules such as proteins, mostly dependent on the duration and intensity of heating. Therefore, proteins are important indicators for the exposure of milk to thermal treatment and consequently possible discriminators between raw and heat-treated milk cheese. Heating of proteins in combination with lactose present in milk generates glycated products following the Maillard reaction. Temperature is considered to have the most effect on the rate of this reaction during food processing and storage (Erbersdobler, 1986). Besides glycation, dephosphorylation is another protein modification that may arise during thermal treatment (Liu et al., 2019). Dephosphorylation has been reported to occur to a limited extent in milk (Lorenzen et al., 2011; Nieuwenhuijse & Van Boekel, 2003). The breakdown of milk proteins into smaller peptides can occur either by enzymatic processes or non-enzymatically through, e.g., heating. Recently, McGrath, Kelly, and Huppertz (2016) reported the heat-induced hydrolysis of sodium caseinate. It was shown that, at very high temperatures, α_{s1}-casein is the most susceptible to the heat-induced hydrolysis and κ-casein is the most thermo-resistant casein.

Liquid chromatography-mass spectrometry (LC-MS) has emerged to a central role in different fields, including food analysis for identification and quantification of proteins in complex matrices due to advantages such as improved accuracy, precision, better selectivity and multiplex peptide/protein analysis in one experiment (Donato, Cacciola, Beccaria, Mondello, & Dugo, 2012). Various authentication methods have been used for protein identification purposes in dairy products, including immunological, spectroscopic, LC-MS-based or DNA-based techniques (Kuckova, Zitkova, Novotny, & Smirnova, 2019). Among these techniques, LC-MS based techniques combined with multivariate techniques are able to accurately predict the levels of milk species adulteration of raw or heat-treated milk chesses and are considered an established technique for protein identification capabilities and verifying milk authenticity.

Different mass spectrometry-based studies have provided information on the occurrence of modified proteins, mostly lactosylation, in milk products after heat treatment (Arena, Renzone, Novi, & Scaloni, 2011; Losito, Carbonara, Monaci, & Palmisano, 2007; Meltretter, Wust, & Pischetsrieder, 2014; Siciliano, Rega, Amoresano, & Pucci, 2000). Bottom-up proteomics is a common method to identify proteins and characterise their amino acid sequences and post-translational modifications (PTMs), using proteolytic digestion of proteins prior to liquid chromatography with tandem mass spectrometry (LC-MS/MS). Several articles have described bottom-up proteomics approaches for the characterisation of (heat-dependent) protein modifications. Galvani, Hamdan, and Righetti (2000, 2001) were able to identify lactose-conjugates of caseins and lactoglobulin with two-dimensional electrophoresis (2-DE) and matrix-assisted laser desorption/ionisation with time-of-flight MS (MALDI-TOF MS). Holland, Gupta, Deeth, and Alewood (2011, 2012) and Marvin, Parisod, Fay, and Guy (2002) applied the same approach to investigate milk protein modifications in relation to heat. Several authors used in-solution digestion combined with LC-MS/MS analysis (O'Donnell, Holland, Deeth, & Alewood, 2004; Kuckova et al., 2019; Picariello, Ferranti, Mamone, Roepstorff, & Addeo, 2008; Poulsen, Jensen, & Larsen, 2016; Rauh et al., 2015).

Another approach is top-down proteomics, which analyses endogenous proteins by tandem MS and retains the endogenous protein mass information, enabling the identification and characterisation of novel proteoforms and quantification of PTMs. The main advantages of the top-down approach include the ability to detect degradation products, sequence variants, and combinations of post-translational modifications. Top-down proteomics is therefore a highly suitable method to characterise the complex mixture of modified caseins in cheese and to detect differences between raw and heat-treated milk cheeses. Furthermore, a top-down approach is less time-consuming compared with the bottom-up approach, since no previous enzymatic cleavage of the proteins is required before sample analysis.

Top-down proteomics research for milk-related products is still in its infancy. This approach was used to study κ-caseinoglycomacro peptide forms, a protein produced during cheese-making (Guerrero, Lerno, Barile, & Lebrilla, 2015). Cunsolo, Muccilli, Saletti, and Foti (2011) and Vincent, Elkins, Condina, Ezernieksn, & Rochfort (2016) have shown that protein modification forms, including advanced glycation end products (AGE), can be identified with this approach in milk. As already mentioned above, several methods have been developed for the detection of raw and heat-treated milk. Nevertheless, the development of robust methods for the detection and discrimination of raw and heat-treated milk in cheese is more challenging.

In this study, for the first time is described a proof of principle method to analyse endogenous cheese proteins with LC-MS in combination with multivariate analysis that can be used for the authenticity of 'Boerenkaas'. Subsequently, top-down methods were applied to identify discriminant endogenous proteins/peptides.

2. Material and methods

2.1. Reagents

Guanidine chloride, dithiothreitol (DTT), tris(hydroxymethyl) aminomethane (Tris), sodium citrate, acetic acid, trifluoroacetic acid (TFA), iodoacetamide (IAA) and formic acid (FA) were obtained from Sigma Aldrich (Zwijndrecht, the Netherlands). Water and acetonitrile were obtained from Biosolve (Valkenswaard, the Netherlands). α_{S1}-Casein and β-casein standards were obtained from Protea Biosciences (Morgantown, WV, USA), while the κ-casein standard and trypsin were obtained from Sigma Aldrich.

2.2. Samples

Raw milk and heat-treated milk cheese samples were obtained from COKZ (Netherlands Controlling Authority for Milk and Milk Products) and were randomly sampled from different cheese-making farms in the Netherlands, where each farm applied their own process to produce cheese. The sample-set consisted of 23 raw milk cheeses and eight heat-treated (pasteurised) milk cheese samples. All cheeses were ten days matured. To confirm the identity (raw milk or heat-treated milk) of these cheese samples, they were also tested with two additional methods for determination of the authenticity of 'Boerenkaas'. Phosphatase activity was determined for all samples according to ISO 11816−2:2016. Volatile fingerprints were determined using PTR-MS, and these profiles were converted into class probability scores for each sample according to the multivariate model as previously described (Alewijn et al., 2018).

2.3. Experimental set-up

Samples were divided into two groups; the main group was used as a discovery dataset to identify possible biomarkers. The discovery dataset consisted of 19 raw milk and 6 heat-treated milk cheese samples. All samples included in this set were prepared in duplicate (sample replicates), and one of the duplicates was injected twice (analytical duplicate). The validation dataset consisted of 9 raw milk and 8 heat-treated milk cheese samples, from

which 5 raw milk and 6 heat-treated milk cheese samples were randomly selected from the discovery dataset to determine the reproducibility of the sample clean-up. All samples in the latter series were analysed without replicates and analytical duplicates.

2.4. Sample preparation

Cheese samples were prepared by weighing approximately 3 g in a polypropylene tube. An amount of 8 mL 6 M guanidine-HCl with 0.1 M Tris/5.37 mM sodium citrate/20 mM DTT was added. The mixtures were then incubated and tumble-shaken at room temperature for 90 min, upon which they were centrifuged at 2200×*g* for 10 min. An amount of 100 μL of each sample was put into vials and 10 μL acetic acid was added to each sample. Subsequently, samples were diluted by adding 900 μL 6 M Guanidine-HCl with 0.1 M Tris/5.37 mM sodium citrate/20 mM DTT. They were stored in a freezer at −20 °C until analysis.

2.5. LC-MS equipment

All samples were measured on an HPLC system (Dionex Ultimate 3000 RSLC) equipped with an NCS-3500RS binary nanoLC pump with an integrated ternary gradient loading pump and column oven, a WPS-3000TPL thermostat autosampler and VWD-3400RS variable wavelength detector (Dionex Softron GmbH, Germering, Germany). Mass spectrometric detection was performed on a Q Exactive mass spectrometer (Thermo Scientific, Waltham, MA, USA) operated in the positive ion mode using an electrospray ionisation (ESI) source. Xcalibur 3.0.63 software was used for data acquisition, processing and analysis.

2.6. LC-MS/(MS) analysis of endogenous cheese proteins

2.6.1. Chromatography for endogenous cheese protein analysis

Chromatographic separation of the prepared cheese samples was achieved on an Aeris Widepore XB-C8 column (3.6 μm; 200 Å; 2.1 × 250 mm) from Phenomenex (Maarssen, the Netherlands) at 60 °C. All samples were analysed in random order, and the injected sample volume was 5 μL. The mobile phase consisted of solvents A (water with 1% formic acid, v/v) and B (acetonitrile with 1% formic acid, v/v). A flow rate of 0.2 mL min-1 was used with a gradient starting at 80% A to 72% A in 2.5 min. At 30 min, A was decreased to 60%. The column was then flushed until 37.5 min with 99% B upon which the eluent was changed to 80% A for equilibration of the column until 45 min.

2.6.2. LC-MS

Full scan MS was performed to identify differences between raw milk cheese and heat-treated milk cheese. The scan range was 700–2500 *m/z* and the in-source CID was set to 20 eV. Five microscans were recorded, the resolution was set at 140,000, the AGC target to 3×10^6, the inject time was 200 ms, S-lens RF level was 80, and the capillary temperature was 325 °C.

2.6.3. Data analysis

Datafiles were converted to mzXML format with MSConvert 3.0 (https://proteowizard.sourceforge.net/download.html). The data for each sample that were not between a retention time of 6 and 28 min were excluded, since all peptides and proteins are eluting from the column in this time range. Subsequently, XCMSonline (https://xcmsonline.scripps.edu) was used for peak picking and alignment. Output files were exported from XCMSonline to Excel and further processed as described in Section 2.6.4.

2.6.4. Multivariate statistics

To perform statistical analysis on the collected data, the results were combined in an Excel sheet containing all variables. These variables consist of all *m/z* values, retention times and peak areas. The Excel sheet generated under data-analysis was imported into Simca 15 software (Umetrics). The data were log10-transformed and Pareto scaled. The data were first examined by principal component analysis (PCA) to assess the analytical precision of the dataset. Orthogonal partial least square discriminant analysis (OPLS-DA) was used for the determination of the compounds that are responsible for the separation of the heat-treated milk and raw milk cheese. The OPLS-DA model was validated through cross-validation, which is described by R^2, Q^2 values. The R^2 value is the proportion of variance in the data explained by the model and indicates the goodness of the fit. The Q^2 parameter is the proportion of variance in the data predictable by the model and indicates predictability. The OPLS-DA model was validated using the internal cross-validation default method (7-fold). Similar observations in the same group are kept together. Further evaluation of the model was performed using a permutation test (100 permutations). A permutation test randomly swaps the identity of samples and re-models the data. A permutation test gives lower R^2 and Q^2 scores on a model not suffering from overfitting. These values should always be lower than the original values. To determine the biomarkers that are responsible for the separation of the groups in the OPLS-DA model, an S-plot was used. The biomarkers of interest were found by the selection of the upper right and lower left biomarkers projected in the S-plot. These were confirmed using the variable importance in projection scores (VIP).

2.6.5. LC-MS/MS (top-down proteomics)

Using the same chromatographic separation as described above, top-down MS was carried out to identify the most discriminating features according to the multivariate data analysis. An inclusion list of these peaks was created, and the top 5 most intense signals were fragmented with NCE set at 30 eV, making use of the active exclusion. The scan range was 700–2500 *m/z*, the in-source CID was set at 20 eV, the resolution at 70,000 for MS and 120,000 for MS/MS. The AGC target was to 5×10^5 for both MS and MS/MS, the inject time was 100 ms for MS and 250 ms for MS/MS. Five microscans were recorded. S-lens RF level was 80 and the capillary temperature was 325 °C. Results were analysed with Prosight Lite 1.4.7 (http://prosightlite.northwestern.edu/).

2.6.6. Top-down proteomics - data analysis

RAW datafiles were deconvoluted with Xtract (part of the Xcalibur software) to identify the precursor mass of the proteins. The parameters used were: mass range 500–2500 *m/z*, resolution 10,000 at 400 *m/z* and S/N threshold was 10. The *m/z* list of the fragments of the precursors subjected to identification was imported into Prosight Lite 1.4.7 (http://prosightlite.northwestern.edu/). FASTA files of bovine milk proteins (Uniprot) were used for protein identification. The average masses deconvoluted with Xtract were used as precursor mass parameters. The precursor mass tolerance with which it was searched in Prosight Lite was 20 ppm.

2.7. LC-MS/MS analysis of tryptic digested proteins

To confirm the identified biomarkers by the top-down approach, samples were fractionated and digested the cheese samples with trypsin and applied bottom-up methods as described below.

2.7.1. Sample fractionation

Two samples from each group were selected and fractionated with half-minute collections by hand using the chromatography method as described in section 2.6.1. Bottom-up analysis (described below) were performed on each fraction. Data obtained from these experiments were used for protein identification.

2.7.2. Sample digestion

An amount of 5 μL 100 mM DTT was added to 100 μL cheese extract (see sample preparation in section 2.4) of each collected fraction and heated for 30 min at 60 °C. Samples were then cooled for 10 min and 5 μL 200 mM IAA was added before incubation in the dark at room temperature for 30 min. Samples were then let in the light for 15 min and quenched with 2 μL 100 mM DTT. Samples were digested overnight with 2 μL 1 ug μL^{-1} trypsin at 37 °C. To terminate digestion, 5 μL FA was added. Samples were stored in a freezer at −20 °C until analysis.

2.7.3. Chromatography for peptide analysis

The digested samples were injected into an Aeris peptide XB-C18 column (1.7 μm; 100 Å; 150 × 2.1 mm) from Phenomenex (Maarssen, the Netherlands) at 50 °C. The injected sample volume was 5 μL. The mobile phase consisted of solvents A: water with 0.1% formic acid (v/v) and B: acetonitrile with 0.1% formic acid (v/v). A flow rate of 0.25 mL min^{-1} was used with a gradient starting with equilibration of 5 min at 95% A, followed by a decrease to 65% A in 30 min. Subsequently, a further decrease to 50% A in 5 min was applied. The column was then flushed for 4 min with 90% B, upon which the eluent was changed to 95% A again for equilibration of the column until 48 min.

2.7.4. LC-MS/MS (bottom-up proteomics)

A scan range of 375−1600 *m/z* was applied in MS and 200−2000 *m/z* in MS/MS. Data were collected in the profile mode. The resolution was set at 70,000 for MS and 17,500 for MS/MS. The AGC target was to 3×10^6 for MS and to 5×10^4 for MS/MS, the inject time was 10 ms for MS and 100 ms for MS/MS. NCE was set to 25 eV.

2.7.5. Bottom-up proteomics - data analysis

RAW datafiles were imported into Mascot 2.5.1. A FASTA bovine protein database (Uniprot) was used for peptide identification. Search parameters were: a precursor mass tolerance of 10 ppm, a fragment mass tolerance of 0.05 Da, a maximum number of missed cleavages of 1, common fixed and variable modifications (carboxyamidation and oxidation, respectively) and FDR of 1%.

3. Results and discussion

3.1. Endogenous cheese protein analysis

In this study, LC-MS methods (as described in Section 2) were developed for analysis of endogenous cheese proteins from raw and heat-treated milk cheese samples. However, both a phosphatase and a PTR-MS fingerprint analyses (section 2.2) were a priori performed to confirm that the heat-treatment status of the used samples was correct (results not shown). Full scan MS was applied on extracted cheese samples with the aim to detect endogenous protein masses that discriminate between raw milk cheese and heat-treated milk cheese. Most of the observed species have low molecular masses (<15 kDa). The intensity of the higher molecular species is much lower, a reason for the suppression of those compounds could be related to the analysis of the relatively complex sample in a normal mode of the Q Exactive. The presence of low molecular masses in the cheese samples indicates that some of the milk proteins are either fragmented in the mass spectrometer or are degraded during sample preparation or already in the cheese-making process. A number of the detected low molecular masses in the cheese samples are most likely an effect of heating and proteolysis by different types of enzymes. Plasmin is one of the most active proteolytic agents in cheeses from milk heated at high temperatures. Plasmin hydrolyses α_{S1}-, α_{S2}-, β- and κ-caseins (Taivosalo et al., 2017; Walstra & Jenness, 1984). Proteolysis is also known to be an important biochemical process during cheese ripening, resulting in a variety of small and large peptides (Feijoo-Siota et al., 2014). Enzymes such as chymosin and other starter microbial proteases used in the cheese-making production play an important role in the proteolysis of milk. κ-casein is hydrolysed by chymosin during the cheese production into *para*-κ-casein and glycomacropeptides (GMP).

Multivariate analysis was subsequently used on the LC-MS data to detect endogenous discriminatory protein biomarkers between 'Boerenkaas' and heat-treated milk cheese. For the evaluation of the analytical stability, PCA modelling was performed using the discovery dataset, consisting of 19 raw milk and 6 heat-treated milk cheese samples (Fig. 1). These results can only be useful if both the analytical duplicates (depicted with circle symbols) and the independent sample replicate (depicted with circle* symbol combination) cluster together in the PCA-score plot. The observed tight clustering of the triplicates (the same colour) reveals strong analytical repeatability, further highlighted by the analytical duplicates, showing less spread than the between-sample variance. For three samples (dark purple symbols on the right side, light green symbols in the upper left side and light purple symbols in the middle left of Fig. 1), there is a clear distance between the replicates, indicating that there is some variation in the sample processing for these samples. It is unclear if this variation is due to sample inhomogeneity, sample clean-up or differences in LC-MS measurement. Since there are only three samples out of 25 that show this variability, the developed method is robust and further analyses on the data can be performed.

Next, an OPLS-DA model was built to find biomarkers to discriminate 'Boerenkaas' cheese made from raw or heat-treated milk using the discovery dataset (full circle symbols). The graphical representation of the OPLS-DA analysis is shown in Fig. 2, showing a clear separation between the two groups. The $R^2 = 0.99$ (measurement of fit) and $Q^2 = 0.98$ (prediction of the model according to cross-validation) were both higher than > 0.9, indicating that the model is capable of predicting the right class. The OPLS-DA permutation test was good (Supplementary material Fig. S1) and demonstrated that the model is unlikely to be based on chance.

The raw milk cheese and the heat-treated milk cheese samples are clearly separated from each other on the x-axis in Fig. 2a. Raw milk cheese samples are more scattered than heated milk cheeses, indicating that this group is more heterogeneous. All samples used for the model building were also correctly classified in the OPLS-DA model. Overall, it can be concluded that the OPLS-DA model is robust and passes all the internal model validation checks. To further validate the model, a second dataset described in Section 2.3, the validation dataset, was projected in the OPLS-DA model. The projection of the validation dataset in the discovery dataset is shown in Fig. 2a. All projected samples were classified in the correct sample group, further validating the model. This result confirms that the combination of analytical results and statistical analysis are robust and that the OPLS-DA model can be used for the discrimination of the two groups of cheese samples.

To determine which features are important for the separation of the two groups, SIMCA's loadings S-plot™ was constructed using the validated OPLS-DA model (Fig. 2b). The S-plot features projected in the upper right and lower left selections of the S-plot were selected and considered as the most robust discriminating

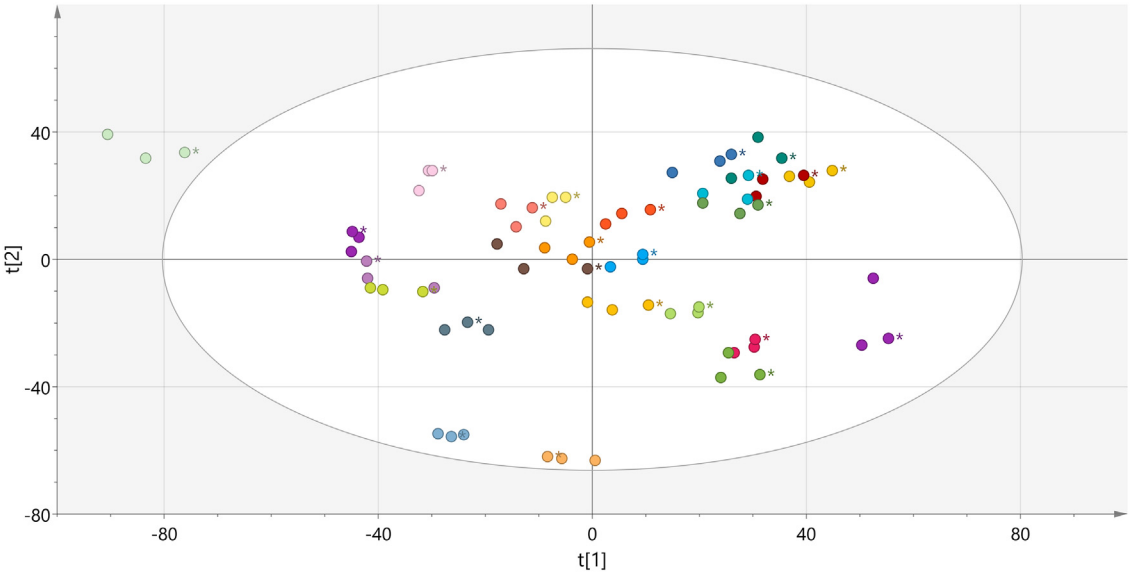

Fig. 1. PCA-score plot of the discovery dataset for determination of the analytical performance. The triplicates belonging to one cheese sample have the same colour. The replicate samples (which are not injected in duplicate) are denoted with the circle* symbol. t represents the fraction of the total variation that can be explained by each component: t [1] = 0.244 and t[2] = 0.166. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

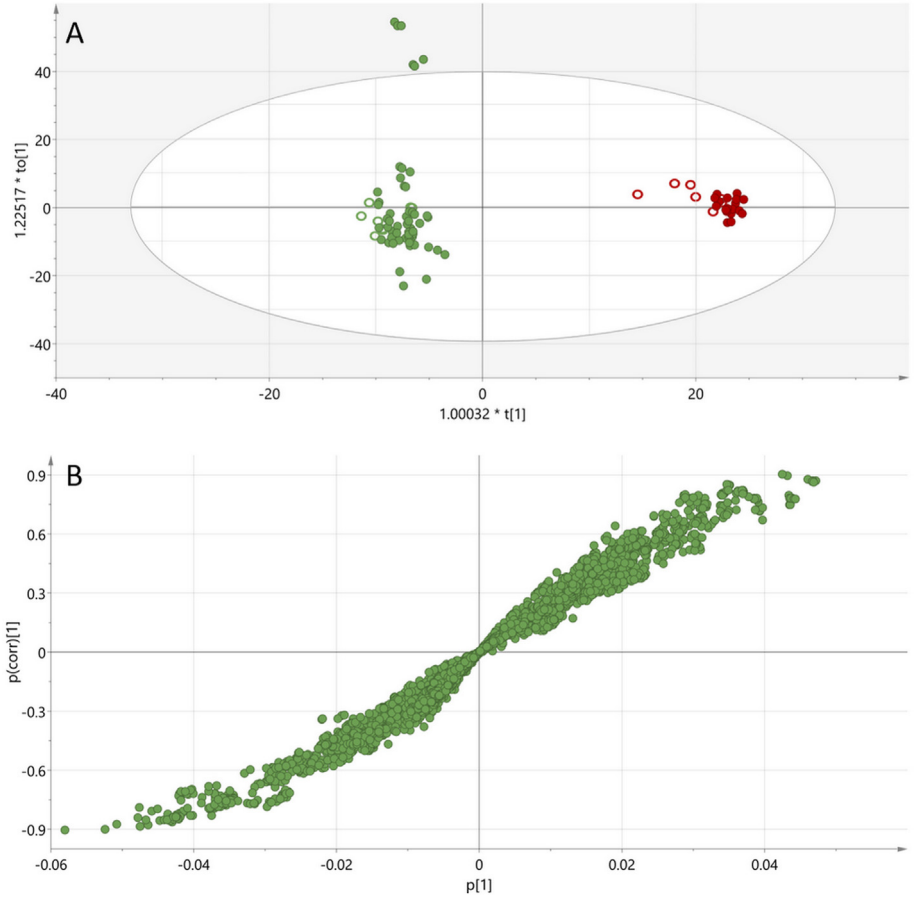

Fig. 2. (a) OPLS-DA score plot and (b) its corresponding S-plot of the discovery dataset: raw milk cheese (green) and the heat-treated milk cheese (red) samples. Samples of the discovery dataset are denoted with the full circle symbols and the projected samples of the validation dataset in the model of discovery dataset are denoted with the open circle symbols. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

biomarkers. The importance of each biomarker chosen for the separation of the two groups was assessed using the variable importance for the projection (VIP) plot (Supplementary material Fig. S2).

The peak-area in each sample for two of the most relevant biomarkers for the separation of both groups, *m/z* 724.5741 (z = 6, RT 10.3 min) and 1551.9423 (z = 7, RT 11.3 min), are shown in Fig. 3a and b. Although the samples were randomly analysed, in Fig. 3 the analytical duplicates, followed by the independent sample replicates of each cheese sample, are grouped together.

The peak-area of the biomarkers in the heat-treated milk cheese are significantly decreased compared with raw milk cheese, clearly demonstrating that these biomarkers are discriminating in determining the authenticity of the cheese. The threshold level that determines if a cheese is made from heat-treated or raw milk (see the blue-dashed line in Fig. 3) is set in the middle of the peak-area value of the lowest raw milk cheese sample and the highest heat-treated milk cheese sample. By use of this tentative threshold, all samples can be correctly classified.

3.2. Identification of the two proposed biomarkers with top-down proteomics

For the identification of these two biomarkers, top-down LC-MS/MS analysis was performed. Based on the fragmentation of the masses of interest, *m/z* 724.5741 and *m/z* 1551.9423, C-terminal truncated forms of κ-casein and α_{S2}-casein were identified. FASTA files of whey and casein protein of bovine milk proteins (Uniprot) were used for protein identification. The MS/MS spectrum of one of the discriminant biomarkers, *m/z* 1551.9423, detected in raw milk cheese, and the identified fragments are shown in Fig. 4. Three fragments originating from the C-terminus were matched in the MS/MS spectrum of *m/z* 724.5741 and six fragments for *m/z* 1551.9423, respectively. The reason why more fragments have not been matched is most likely due to the variety of proteoforms of these two proteins that can be found. Both κ-casein and αs_2-casein are existing in different forms since they have multiple amino acid sites that can contain PTMs, such as glycosylation and phosphorylation. Furthermore, other unknown modifications could occur during sample processing. Since the experimental mass of the truncated proteins could not be matched with the calculated theoretical masses of different combinations of modified C-terminal truncated forms, the C-terminal truncated forms could not be fully characterised. Therefore, both could not be identified both, the exact cleavage site and/or possible amino acid modifications of these proteins.

From the major milk proteins, whey proteins are mainly eliminated in the whey at draining (Bramanti, Sortino, Onor, Beni, & Raspi, 2003; Montagne et al., 1995; Muller-Renaud, Dupont, & Dulia, 2004) whereas the caseins transfer to curd and cheese. Therefore, it is not surprising that two of the most discriminant biomarkers between raw and heat-treated milk are truncated

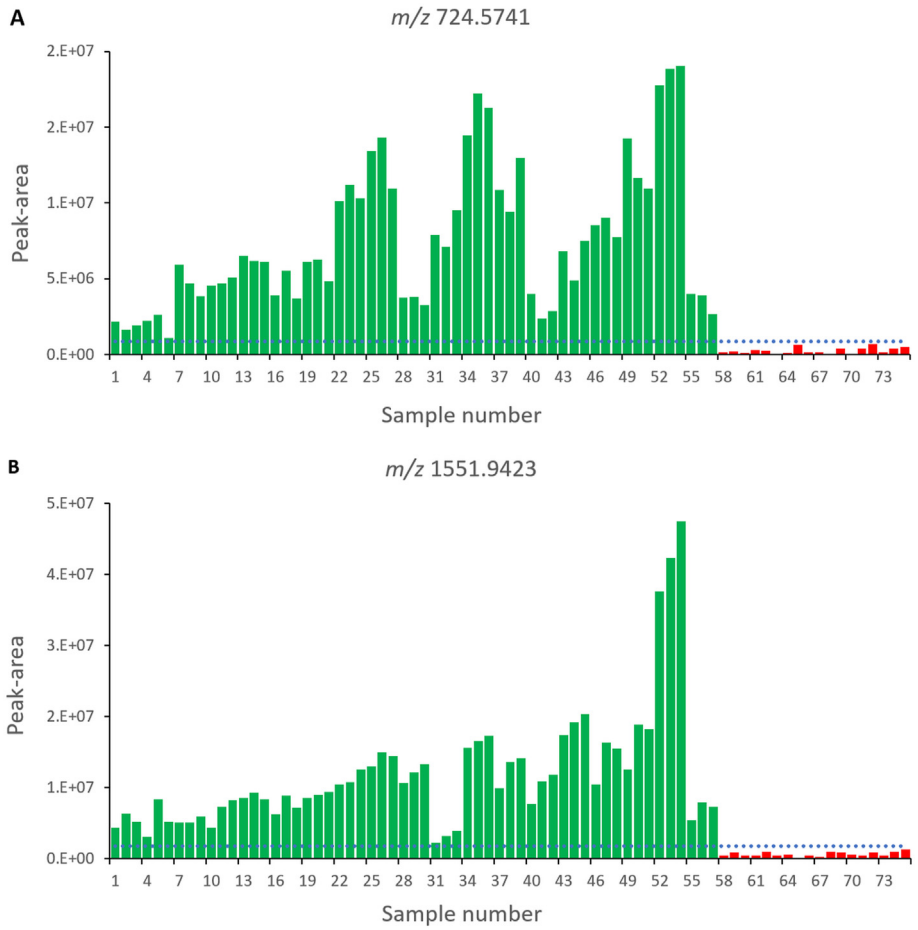

Fig. 3. Peak-areas in each sample for two selected potential biomarkers, (a) *m/z* 724.5741 and (b) *m/z* 1551.9423, respectively, in raw-milk (green) and heat-treated milk cheese (red) samples. Blue-dashed lines represent threshold levels to determine if a cheese is made from heat-treated or raw milk. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

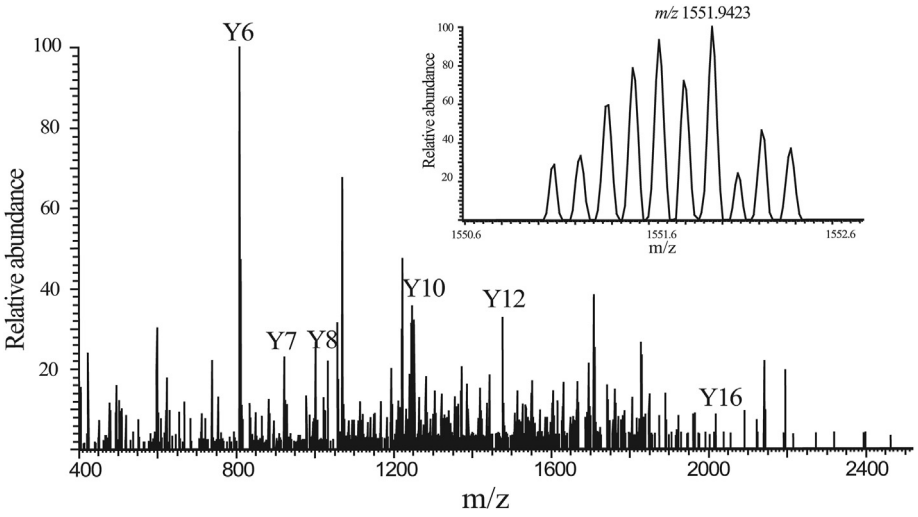

Fig. 4. Top-down spectrum of *m/z* 1551.9423 (z = 7). The identified fragments are annotated in the figure. In the right-upper corner, the spectrum of the parent ion is shown.

forms of two caseins (κ- and α_{S2}-casein). Truncation of the caseins could be the result of protease activity. In the heat-treated milk cheeses, these two truncated forms are detected at very low levels or not detected at all, probably being further hydrolysed by enzymes or post-translationally modified with varying levels of phosphorylation, glycosylation and oxidation of cysteine to disulphide bonds (Vincent, Elkins, Condina, Ezernieks, & Rochfort, 2016). Aggregation of whey proteins such as β-lactoglobulin with non-whey proteins such as κ-casein upon heat treatment, could reduce the hydrolysis of κ-casein in heat-treated milk cheese, this being another reason for the absence of the identified truncated form of κ-casein in heat-treated (Wijayanti, Bansal, & Death, 2014).

3.3. Confirmation of the biomarkers with bottom-up proteomics

To validate the identification of the truncated forms of κ-casein and αs_2-casein by the top-down approach, a bottom-up method was applied (see Section 2.7.). The fractions containing the masses of interest were digested with trypsin and analysed with the bottom-up approach. κ-casein was identified as the main protein in the fraction at 10.3 min, with 15% sequence coverage and α_{s2}-casein was identified as the main protein in the fraction at 11.3 min with 53% sequence coverage. The bottom-up approach confirms the presence of κ-casein and α_{S2}-casein in the corresponding fractions. The identified peptides are included in Supplementary material Table S1.

4. Conclusions

The developed extraction and endogenous cheese LC-MS method enable the measurement of endogenous proteins/peptides from cheese. This method was applied to differentiate between cheeses made from raw and heat-treated milk. An OPLS-DA model was built based on the proteome data. The OPLS-DA model was successfully validated by testing it on a second dataset. From the S-plot and VIP plot, various *m/z* values were detected as discriminatory masses between the two groups. The two most discriminating biomarkers were identified by both top-down approach and are truncated forms of κ-casein and α_{S2}-casein. The identification of both proteins was confirmed by bottom-up approach. For future research, it is recommended to measure and analyse a larger dataset with the LC-MS method described in this paper to determine the robustness of both biomarkers, including for cheeses made from milk with mild heat-treatments (thermisation).

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 thank Centraal Orgaan voor Kwaliteitsaangeleheden in de Zuivel (COKZ) for providing the cheese samples. The authors acknowledge the financial support by Ministry of Agriculture, Nature and Food Quality.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.idairyj.2021.104990.

References

Alewijn, M., Wehrens, R., & van Ruth, S. M. (2018). Robust detection methodology of milk heat treatment in cheese based on volatile profile fingerprinting. *International Dairy Journal, 85*, 211–218.

Arena, S., Renzone, G., Novi, G., & Scaloni, A. (2011). Redox proteomics of fat globules unveils broad protein lactosylation and compositional changes in milk samples subjected to various technological procedures. *Journal of Proteomics, 74*, 2453–2475.

Bergamaschi, M., Cecchinato, A., Biasioli, F., Gasperi, F., Martin, B., & Bittante, G. (2016). From cow to cheese: Genetic parameters of the flavour fingerprint of cheese investigated by direct-injection mass spectrometry (PTR-ToF-MS). *Genetics Selection Evolution, 48*, Article 89.

Birlouez-Aragon, I., Sabat, P., & Gouti, N. (2002). A new method for discriminating milk heat treatment. *International Dairy Journal, 12*, 59–67.

Boitz, L. I., Fiechter, G., Seifried, R. K., & Mayer, H. K. (2015). A novel ultra-high-performance liquid chromatography method for the rapid determination of b-lactoglobulin as heat load indicator in commercial milk samples. *Journal of Chromatography A, 1386*, 98–102.

Bramanti, E., Sortino, C., Onor, M., Beni, & F., & Raspi, G. (2003). Separation and determination of denatured αs1-, αs2-, β- and κ-caseins by hydrophobic interaction chromatography in cows', ewes' and goats' milk, milk mixtures and cheeses. *Journal of Chromatography A, 994*, 59–74.

Claeys, W. L., Van Loey, A. M., & Hendrickx, M. E. (2002). Intrinsic time temperature integrators for heat treatment of milk. *Trends in Food Science & Technology, 13*, 293–311.

Cunsolo, V., Muccilli, V., Saletti, R., & Foti, S. (2011). Applications of mass spectrometry techniques in the investigation of milk proteome. *European Journal of Mass Spectrometry, 17*, 305–320.
Donato, P., Cacciola, F., Beccaria, M., Mondello, L., & Dugo, P. (2012). Advances in LC-MS for food analysis. *LC-GC Europe, 25*, 15–24.
Ebner, J., Baum, F., & Pischetsrieder, M. (2016). Identification of sixteen peptides reflecting heat and/or storage induced processes by profiling of commercial milk samples. *Journal of Proteomics, 147*, 66–75.
Erbersdobler, H. F. (1986). Twenty years of Furosine-Better knowledge about the biological significance of Maillard reaction in food and nutrition. In M. Fujimaki, M. Namiki, & H. Kato (Eds.), *Amino-carbonyl reaction in food and biological systems* (pp. 481–491). Amsterdam, The Netherlands: Elsevier.
Feijoo-Siota, L., Blasco, L., Rodriguez-Rama, J. L., Barros-Velazquez, J., de Miguel, T., Sanchez-Perez, A. G., et al. (2014). Recent patents on microbial proteases for the dairy industry. *Recent Advances in DNA & Gene Sequences, 8*, 44–55.
Galvani, M., Hamdan, M., & Righetti, P. G. (2000). Two-dimensional gel electrophoresis/matrix-assisted laser desorption/ionisation mass spectrometry of a milk powder. *Rapid Communications in Mass Spectrometry, 14*, 1889–1897.
Galvani, M., Hamdan, M., & Righetti, P. G. (2001). Two-dimensional gel electrophoresis/matrix-assisted laser desorption/ionisation mass spectrometry of commercial bovine milk. *Rapid Communications in Mass Spectrometry, 15*, 258–264.
Gasperi, F., Gallerani, G., Boschetti, A., Biasioli, F., Monetti, A., Boscaini, E., et al. (2001). The mozzarella cheese flavour profile: A comparison between judge panel analysis and proton transfer reaction mass spectrometry. *Journal of the Science of Food and Agriculture, 81*, 357–363.
Griffiths, M. W. (1986). Use of milk enzymes as indices of heat treatment. *Journal of Food Protection, 49*, 696–705.
Guerrero, A., Lerno, L., Barile, D., & Lebrilla, C. B. (2015). Top-down analysis of highly post-translationally modified peptides by Fourier transform ion cyclotron resonance mass spectrometry. *Journal of the American Society for Mass Spectrometry, 26*, 453–459.
Holland, J. W., Gupta, R., Deeth, H. C., & Alewood, P. F. (2011). Proteomic analysis of temperature-dependent changes in UHT milk. *Journal of Agricultural and Food Chemistry, 59*, 1837–1846.
Holland, J. W., Gupta, R., Deeth, H. C., & Alewood, P. F. (2012). UHT milk contains multiple forms of αS1-casein that undergo degradative changes during storage. *Food Chemistry, 133*, 689–696.
Johnson, P., Philo, M., Watson, A., & Mills, E. C. (2011). Rapid fingerprinting of milk thermal processing history by intact protein mass spectrometry with non-denaturing chromatography. *Journal of Agricultural and Food Chemistry, 59*, 12420–12427.
Kuckova, S., Zitkova, K., Novotny, O., & Smirnova, T. (2019). Verification of cheeses authenticity by mass spectrometry. *Separation Science, 42*, 3487–3496.
Lin, S., Sun, J., Cao, D., Cao, J., & Jiang, W. (2010). Distinction of different heat-treated bovine milks by native-PAGE fingerprinting of their whey proteins. *Food Chemistry, 121*, 803–808.
Liu, H., Grosvenor, A. J., Li, X., Wang, X.-L., Ma, Y., Clerens, S., et al. (2019). Changes in milk protein interactions and associated molecular modification resulting from thermal treatments and storage. *Journal of Food Science, 84*, 1737–1745.
Lorenzen, P. C., Clawin-Raedecker, I., Einhoff, K., Hammer, P., Hartmann, R., Hoffmann, W., et al. (2011). A survey of the quality of extended shelf life (ESL) milk in relation to HTST and UHT milk. *International Journal of Dairy Technology, 64*, 166–178.
Losito, I., Carbonara, T., Monaci, L., & Palmisano, F. (2007). Evaluation of the thermal history of bovine milk from the lactosylation of whey proteins: An investigation by liquid chromatography-electrospray ionization mass spectrometry. *Analytical and Bioanalytical Chemistry, 389*, 2065–2074.
Marvin, L. F., Parisod, V., Fay, L. B., & Guy, P. A. (2002). Characterisation of lactosylated proteins of infant formula powders using two-dimensional gel electrophoresis and nanoelectrspray mass spectrometry. *Electrophoresis, 23*, 2505–2512.
McGrath, B. A., Kelly, A. L., & Huppertz, T. (2016). Proteomic characterisation of heat-induced hydrolysis of sodium caseinate. *International Journal of Dairy, 53*, 51–59.
McKellar, R., Emmons, D., & Farber, J. (1991). Gamma-glutamyl transpeptidase in milk and butter as an indicator of heat treatment. *International Dairy Journal, 1*, 241–251.
Meltretter, J., Wust, J., & Pischetsrieder, M. (2014). Modified peptides as indicators for thermal and nonthermal reactions in processed milk. *Journal of Agricultural and Food Chemistry, 62*, 903–10915.
Montagne, P., Cuilliere, M. L., Marchal, E., El-Bari, N., Montagne, M., Benali, et al. (1995). Application of microparticle-enhanced nephelometric immunoassays of alpha, beta and kappa caseins to evaluate quality of milk, from the production to the valorization in cheese industry. *Le Lait, 75*, 211–237.
Muller-Renaud, S., Dupont, D., & Dulieu, P. (2004). Quantification of κ-casein in milk by an optical immunosensor. *Food and Agricultural Immunology, 15*, 265–277.
Nieuwenhuijse, J. A., & Van Boekel, M. A. J. S. (2003). Protein stability in sterilized milk and milk products. In P. F. Fox, & P. L. H. McSweeney (Eds.), *Advanced dairy chemistry* (2nd ed., Vol. 1, pp. 947–974). New York, NY, USA: Springer US. Proteins.
O'Donnell, R., Holland, J. W., Deeth, H. C., & Alewood, P. F. (2004). Milk proteomics. *International Dairy Journal, 14*, 1013–1023.
Picariello, G., Ferranti, P., Mamone, G., Roepstorff, P., & Addeo, F. (2008). Identification of N-linked glycoproteins in human milk by hydrophilic interaction liquid chromatography and mass spectrometry. *Proteomics, 8*, 3833–3847.
Poulsen, N. A., Jensen, H. B., & Larsen, L. B. (2016). Factors influencing degree of glycosylation and phosphorylation of caseins in individual cow milk samples. *Journal of Dairy Science, 99*, 3325–3333.
Rauh, V. M., Johansen, L. B., Bakman, M., Ipsen, R., Paulsson, M., Larsen, L. B., et al. (2015). Protein lactosylation in UHT milk during storage measured by liquid chromatography-mass spectrometry and quantification of furosine. *International Journal of Dairy Technology, 68*, 486–494.
Recio, I., Amigo, L., & Lopez-Fandino, R. (1997). Assessment of the quality of dairy products by capillary electrophoresis of milk proteins. *Journal of Chromatography B: Biomedical Sciences and Applications, 697*, 231–242.
Siciliano, R., Rega, B., Amoresano, A., & Pucci, P. (2000). Modern mass spectrometric methodologies in monitoring milk quality. *Analytical Chemistry, 72*, 408–415.
Taivosalo, A., Krisciunaite, T., Selman, A., Part, N., Stulova, I., & Vilu, R. (2017). Comprehensive analysis of proteolysis during 8 months of ripening of high-cooked Old Saare cheese. *Journal of Dairy Science, 101*, 1–24.
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, 10*, Article 1371.
Walstra, P., & Jenness, R. (1984). *Dairy chemistry and physics*. New York, NY, USA: John Wiley and Sons.
Wijayanti, H. B., Bansal, N., & Deeth, H. C. (2014). Stability of whey proteins during thermal processing: A review. *Comprehensive Reviews in Food Science and Food Safety, 13*, 1235–1251.