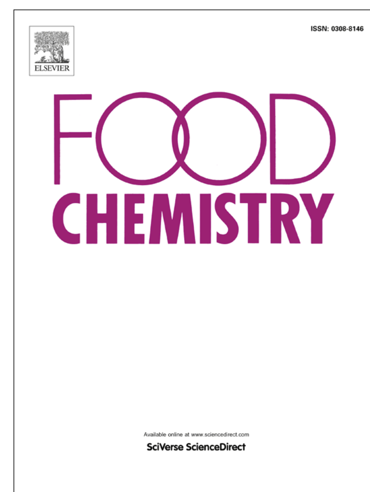
Accepted Manuscript



Proteomic analysis of goat milk kefir: profiling the fermentation-time dependent protein digestion and identification of potential peptides with biological activity

Juan J. Izquierdo-González, Francisco Amil-Ruiz, Sabina Zazzu, Rosa Sánchez-Lucas, Carlos A. Fuentes-Almagro, Manuel J. Rodríguez-Ortega

PII: S0308-8146(19)30991-4
DOI: https://doi.org/10.1016/j.foodchem.2019.05.178
Reference: FOCH 24904

To appear in: *Food Chemistry*

Received Date: 25 March 2019
Revised Date: 22 May 2019
Accepted Date: 24 May 2019

Please cite this article as: Izquierdo-González, J.J., Amil-Ruiz, F., Zazzu, S., Sánchez-Lucas, R., Fuentes-Almagro, C.A., Rodríguez-Ortega, M.J., Proteomic analysis of goat milk kefir: profiling the fermentation-time dependent protein digestion and identification of potential peptides with biological activity, *Food Chemistry* (2019), doi: https://doi.org/10.1016/j.foodchem.2019.05.178

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Proteomic analysis of goat milk kefir: profiling the fermentation-time dependent protein digestion and identification of potential peptides with biological activity

Juan J. Izquierdo-González[1, a], Francisco Amil-Ruiz[2], Sabina Zazzu[1], Rosa Sánchez-Lucas[1], Carlos A. Fuentes-Almagro[3], Manuel J. Rodríguez-Ortega[1] *.

[1]Departamento de Bioquímica y Biología Molecular, Universidad de Córdoba; Campus de Excelencia Internacional CeiA3, Córdoba (Spain).

[2]Unidad de Bioinformática, Servicio Central de Apoyo a la Investigación (SCAI), Universidad de Córdoba, Córdoba (Spain).

[3]Unidad de Proteómica, Servicio Central de Apoyo a la Investigación (SCAI), Universidad de Córdoba, Córdoba (Spain).

[a]Present address: Department of Immunology, University Medicine Greifswald, Sauerbruchstraße DZ7, 17475 Greifswald (Germany).

*Corresponding author:

Manuel J. Rodríguez-Ortega

Address: Departamento de Bioquímica y Biología Molecular, Edificio "Severo Ochoa" planta baja, Campus de Rabanales, Universidad de Córdoba. 14071 Córdoba, Spain.

Tel: +34 957 218519

Fax: +34 957 218856

e-mail: mjrodriguez@uco.es

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Abstract

Kefir is a fermented dairy product, associated to health benefits because of being a probiotic and due to the presence of molecules with biological activity. In this work, we have profiled the peptide composition of goat milk kefir at three different fermentation times using a peptidomics approach, in order to study changes in peptide concentrations and patterns of protein digestion throughout the fermentation time. We identified 2328 unique peptides corresponding to 22 protein annotations, with a maximum of peptides found after 24 h fermentation. We established different digestion patterns according to the nature of the proteins, and quantified the changes in the peptides appearing in all the fermentation times. We also identified 11 peptides that matched exactly to sequences with biological activity in databases, almost all of them belonging to caseins. This is the most comprehensive proteomic analysis of goat milk kefir to date.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

1. Introduction

In the Mediterranean diet, which is recognized as one of the healthiest lifestyles, milk and the derived dairy products like cheese, fermented beverages and others, are an important source of proteins. In addition, they also provide vitamins, minerals and other molecules with important biological activities (Roncada, Piras, Soggiu, Turk, Urbani, & Bonizzi, 2012).

Goat was one of the ruminant species to be earliest domesticated. Goats provide diverse sources of food (meat, milk) in low favourable climatic and environmental conditions, and are resilient animals (Selvaggi, Laudadio, Dario, & Tufarelli, 2014). Although quantitatively less important than bovine livestock nowadays, especially in Europe, goat farming is a major livestock activity in the Mediterranean countries for economic, social and environmental reasons. Actually, Europe possesses around 5% of the World's dairy goat herds, but produces more than 15% of the goat milk in the World (Escareño, Salinas-Gonzalez, Wurzinger, Iniguez, Solkner, & Meza-Herrera, 2013). Particularly, Spain has the second largest number of goats in Europe, being Andalusia the leader region with more than 40% of the Spanish heads of livestock. In Spain, goat milk represents around 14% of total mik production and 25% of milk economic value (FAOSTAT, 2017). In the last decades, goat milk has gained attention because of its less allergenic properties and easier digestibility, compared with cow milk (Clark & Mora Garcia, 2017). Therefore, goat milk and its dairy derivatives are arising as an attractive alternative to the mostly consumed cow milk, with a growing market (Escareño, Salinas-Gonzalez, Wurzinger, Iniguez, Solkner, & Meza-Herrera, 2013).

Kefir is a fermented milk beverage consumed worldwide, whose origin is situated in the Caucasus region and Eastern Europe. Similar in aspect to yogurt, it differs in several features: kefir is made by a symbiosis of lactic acid bacteria and yeasts enclosed in an

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

exopolysaccharide and protein matrix, the kefir grains, which carry out an acidic-alcoholic fermentation. As a result, the fermented product has a tart flavour, and some content of alcohol (around 0.5%). (de Oliveira Leite, Miguel, Peixoto, Rosado, Silva, & Paschoalin, 2013; Rosa, Dias, Grzeskowiak, Reis, Conceicao, & Peluzio, 2017). Kefir consumption has been attributed health benefits, as well as an increase in longevity. These benefits are linked to its use as a source of probiotics, but also to the kefiran, i.e. the grain exopolysaccharide. In addition, other byproducts of the microbial fermentative metabolism have been demonstrated to be beneficial for health. Among them, many peptides resulting from proteolysis of milk proteins have biological activity for a plethora of functions: they have been described as possessing immunomodulating, antimicrobial, antihypertensive, antitumoral or antioxidant properties (Amorim, Coitinho, Dias, Friques, Monteiro, Rezende, et al., 2019; Hayes, Ross, Fitzgerald, & Stanton, 2007; Hayes, Stanton, Fitzgerald, & Ross, 2007; Klippel, Duemke, Leal, Friques, Dantas, Dalvi, et al., 2016; Quiros, Hernandez-Ledesma, Ramos, Amigo, & Recio, 2005). Although bioactive peptides are also present in raw milk, numerous studies have demonstrated that kefir and other fermented products have a notably increased number and variety of this group of molecules, due to the action of microbial proteinases (Amorim, et al., 2019; Dallas, Citerne, Tian, Silva, Kalanetra, Frese, et al., 2016; Ebner, Asci Arslan, Fedorova, Hoffmann, Kucukcetin, & Pischetsrieder, 2015; Piovesana, Capriotti, Cavaliere, La Barbera, Samperi, Zenezini Chiozzi, et al., 2015; Quiros, Hernandez-Ledesma, Ramos, Amigo, & Recio, 2005). Thus, recent works have demonstrated that peptide composition of kefir from bovine milk is rather complex, comprising some hundreds of unique sequences derived from the major milk proteins (Amorim, et al., 2019; Dallas, et al., 2016; Ebner, Asci Arslan, Fedorova, Hoffmann, Kucukcetin, & Pischetsrieder, 2015). However, there are no studies using an untargeted peptidomic approach to define the peptide composition of goat milk kefir. Moreover, there is a lack of studies that characterize the changes of kefir peptide

ACCEPTED MANUSCRIPT

population during the fermentation time and the extent of protein digestion by the action of the microbial proteases.

In the present work, we have studied the peptide profile of goat milk kefir at different fermentation times using a liquid chromatography-tandem mass spectrometry peptidomic approach, to describe the most comprehensive database so far of caprine kefir. Quantitative changes of peptides were observed along fermentation times. We mapped the identified peptides on the protein sequences to elucidate proteolysis patterns depending on digestion time. Finally, we screened for the presence of described and potential peptides with biological activities, which can be used in further studies to explore health benefits derived from goat milk kefir consumption.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

2. Materials and methods

2.1. Reagents and solutions

Acetonitrile (ACN) and formic acid were purchased from Sigma-Aldrich (St. Louis, MO, USA). Solutions containing these reagents were prepared at different % (v/v) in ultrapure water.

2.2. Kefir production

Kefir samples were prepared at three different fermentation times (12, 24, and 36 h) by placing, in sterilized flasks, 5% kefir grains (w/v) in commercial semi-skimmed goat's milk (COVAP, Córdoba, Spain). Fermentation was carried out at 25 ºC in an incubator, in aerobic conditions and without shaking. Non-fermented raw milk (named as 0 h sample) was used as a control. For each sample (0, 12, 24, and 36 h), three independent biological replicates were made. The fermentation product (i.e. kefir) was separated from the kefir grains and coagulated caseins by centrifugation at 5,000 × *g* for 10 min. The supernatants (the kefir samples containing the peptides and other larger polypeptides) were filtered using sterile 0.22-μm pore-size filters (Millipore, Bedford, MA, USA) to remove cells released from the kefir grains. Kefir samples were kept at -20 ºC until further procedure for peptide extraction. Aliquots of raw milk were also frozen to be subject to the same treatment as kefir samples.

2.3. Peptide extraction

Samples kept at -20 ºC were thawed for peptide extraction from complex kefir mixtures. Peptides present in kefir samples and raw milk (control) were separated from non-totally digested proteins and large polypeptides using Amicon ultrafiltration devices (Millipore) with a membrane cut-off of 10 kDa. Four mL of each sample was loaded into the devices and centrifuged at 5,000 × *g* until the whole volume passed throughout the membrane. The

ACCEPTED MANUSCRIPT

peptides contained in the flow-through membrane filtrates were further cleaned and concentrated using Oasis HLB extraction cartridges (Waters, Milford, MA, USA) according to manufacturer's instructions and modified by our research group for peptide cleaning and concentration of samples from bacteria (Rodriguez-Ortega, 2018). Briefly, after conditioning the extraction cartridges with 80% ACN and then with 0.1% formic acid solution, 0.5 mL samples were loaded, and peptides were eluted with increasing concentrations (10, 20, and 50%) of ACN in 0.1% formic acid. Peptide fractions were totally dried with a vacuum concentrator (Eppendorf, Hamburg, Germany), resuspended in 100 µL of 2% ACN/0.1% formic acid and kept at −20 °C until further analysis.

2.4. LC-MS/MS analysis

Peptide separation was performed by nano-LC using a Dionex Ultimate 3000 nano UPLC (Thermo Scientific, San Jose, CA), equipped with a reverse phase C18 75 µm × 50 Acclaim Pepmap column (Thermo Scientific) at 300 nL/min and 40 ºC for a total run time of 85 min. The mix of peptides was previously concentrated and cleaned up on a 300 µm × 5 mm Acclaim Pepmap cartridge (Thermo Scientific) in 2% ACN/0.05% formic acid for 5 min, with a flow of 5 µL/min. Solution A (0.1% formic acid) and Solution B (80% ACN, 0.1% formic acid) were used as mobile phase for the chromatographic separation according to the following elution conditions: 4-35% solution B for 60 min; 35-55% solution B for 3 min; 55-90% solution B for 3 min followed by 8 min washing with 90% solution B, and re-equilibration during 12 min with 4% solution B.

Peptide positive ions eluted from the column were ionized by a nano-electrospray ionization source and analyzed in positive mode on a trihybrid Thermo Orbitrap Fusion (Thermo Scientific) mass spectrometer operating in Top30 Data Dependent Acquisition mode with maximum cycle time of 3 s. Single MS scans of peptide precursors were acquired in a 400-

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

1,500 *m/z* range at 120,000 resolution (at 200 *m/z*) with a 4×10^5 ion count target threshold. For MS/MS, precursor ions were previously isolated in the quadrupole at 1.2 Da, and then CID-fragmented in the ion trap with 35% normalized collision energy. Monoisotopic precursor selection was turned on. Ion trap parameters were: i) the automatic gain control was 2×10^3; ii) the maximum injection time was 300 ms; and iii) only those precursors with charge state 2–5 were sampled for MS/MS. In order to avoid redundant fragmentations a dynamic exclusion time was set to 15 s with a 10-ppm tolerance around the selected precursor and its isotopes.

2.5. Protein identification by database searching

The mass spectrometry raw data were processed using Proteome Discoverer (version 2.1.0.81, Thermo Scientific). Charge state deconvolution and deisotoping were not performed. MS/MS spectra were searched with SEQUEST engine against a database of Uniprot_Capra hircus_Nov2016 (www.uniprot.org), applying the following search parameters: No enzyme was used for theoretical digestion of protein sequences. Methionine oxidation was set as variable modification. A value of 10 ppm was set for mass tolerance of precursor ions, and 0.1 Da tolerance for product ions. Peptide identifications were accepted if they exceeded the filter parameter Xcorr score *versus* charge state with SequestNode Probability Score (+1 = 1.5, +2 = 2.0, +3 = 2.25, +4 = 2.5). Validation of peptide spectral matches (PSM) was done at a 1% FDR using percolator based on q-values. For peptide quantification, precursor ion areas were calculated using Precursor Ion Area Detector and normalized by Total Peptide Amount mode in Proteome Discoverer 2.1. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (Deutsch, Csordas, Sun, Jarnuczak, Perez-Riverol, Ternent, et al., 2017) via the PRIDE (Vizcaino, Csordas, del-Toro, Dianes, Griss, Lavidas, et al., 2016) partner repository with the dataset identifier PXD012142.

ACCEPTED MANUSCRIPT

2.6. Data and statistical analysis

Peptides were considered to be identified in a sample when they were identified in at least two out of the three biological replicates for the given sample. Otherwise, peptides identified only in one biological replicate were not considered as present in the sample(s) and were discarded from the overall count of identified peptides. For further quantitative analysis, means and standard deviations were calculated using an Excel spreadsheet (Microsoft Excel 2011 v14.0.0 for Mac).

Biopeptide searches were done against the Biopep database (http://www.uwm.edu.pl/biochemia/, date 2018.07.30, containing 3669 entries). Quantified peptides from kefir samples were employed to search for biopeptides. For each quantified peptide, a normalization procedure was carried out to allow peptide comparison between samples. Thus, peptide quantification values were corrected to the sum of all peptides quantifications for a given sample. Each biopeptide sequence from the Biopep database was searched within a given sample's peptide list. For every biopeptide perfectly matching a peptide in a given sample, its quantification value was associated to the found biopeptide. In parallel, putative biopeptide precursors were also searched within a given sample's peptide list. Thus, a peptide containing a biopeptide within its sequence was considered as a putative biopeptide precursor. For each biopeptide found with this procedure, its precursor quantification value was associated to the found biopeptide. It could be possible that a particular biopeptide was found several times within a particular peptide sequence. In such cases, all hits were taken into consideration, and a quantification value was calculated for each biopeptide multiplying the number of times that the biopeptide is contained in a precursor by the precursor normalized quantification value. A quantification of the biological activities represented by the biopeptides identified in each sample was done as the sum of the quantification values of all biopeptides contributing to such biological activity, being further

ACCEPTED MANUSCRIPT

normalized by the frequency of the biopeptides contributing to such biological activity in the global Biopep database. For the hierarchical clustering and heatmap analysis, only those peptides identified in all tested samples were retained. Quantification values were z-scored prior to the clustering analysis by Pearson correlation with method "complete" (hclust function in stats package from R).

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

3. Results

3.1. Peptide and protein profiling of kefir at different fermentation times

The general aim of this study consisted of profiling the changes in peptide composition in goat milk kefir throughout fermentation time. For that, we performed an LC-MS/MS analysis of peptide samples in kefir after 12, 24, and 36 h fermentation, using non-fermented raw milk as the control.

In the global analysis of the four samples (i.e. the three fermentation times plus the control), we identified 2328 unique peptides belonging to 32 protein entities, which corresponded to 22 unique protein annotations (Table 1 and Supplementary Dataset 1). We identified the major milk proteins, i.e. caseins (α_{s1}-, α_{s2}-, β-, and κ-casein), as well as the classical major whey proteins (α-lactalbumin and β-lactoglobulin) and other proteins, including the serum amyloid A protein, the butyrophilin subfamily 1 member A1, the glycosylation dependent cell adhesion molecule 1 (GLYCAM-1), lactoperoxidase, osteopontin, and perilipin. Other ten gene products were also found, comprising the Agouti signaling protein, the β-1,4-galactosyltransferase I, the parathyroid hormone-like protein, a fatty acid synthase, a fatty acid binding protein, the fibrinogen α chain, the Cysteine rich secretory protein 3, and three cathelicidin-like gene products: Bac7.5, cathelicidin-2 (Bac5), and MAP28.

Most of the proteins were identified across all the samples, although there were differences: out of the 22 unique protein functions, 7 were missing in the control, 12 were missing in the 12-h fermented sample, and 2 and 3 protein functions were missed in samples corresponding to 24 h and 36 h fermentation, respectively. However, as expected, the differences were quite large at peptide level, both considering the total amount of peptides identified per sample, and for each individual protein throughout the fermentation time: in raw milk we identified 261 unique peptides for the 15 protein functions found in this control group; 327 peptides in the

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

12 h sample; 2004 unique peptides at 24 h; and 1441 peptides in the 36 h sample. Those proteins identified from the highest numbers of peptides were the caseins. Thus, β-casein was identified from 846 peptides; and α_{s1}-, α_{s2}-, and κ-casein were found from 285, 319, and 243 unique peptides respectively, in all the samples. The other proteins, including minority proteins normally associated to whey fraction, were identified from lower numbers of unique peptides, with the exception of the serum amyloid A protein, which was found from a total of 383 peptides. In general, although the numbers of unique peptides were quite variable among proteins within a given sample (e.g. from 1 to 762 in the 24-h fermentation sample), for the major proteins (i.e. caseins and serum amyloid A protein, as well as others like GLYCAM-1 or osteopontin) there was a similar pattern consisting of an increase in identified unique peptides from raw, non-fermented milk to 24 h fermentation, and then a decrease at 36 h. Although in general the peak of identified unique peptides for each protein was found at 24 h fermentation, the aforementioned pattern varied for other minor proteins.

Figure 1 shows the intersection diagram of the unique peptides identified in the four samples. After 24 h fermentation, we found the highest number of exclusive peptides in a sample. Out of the 2328 peptides identified, 715 (30.7%) were found only in the 24 h sample. At 36 h fermentation, 228 unique peptides (9.8%) were exclusive in this sample. Only 9 peptides were found exclusively in the kefir after 12 h fermentation. Interestingly, 81 peptides (3.5% of total) were exclusive in raw milk. The samples corresponding to the two last fermentation times (i.e. 24 and 36 h) shared 882 unique peptides (37.9%), a number significantly much higher than for the rest of kefir sample pairs (i.e. 12 and 24 h; 12 and 36 h). The raw milk and the 12 h sample shared only 7 unique peptides, and 8 when considering also the 24 h sample. Of note, 56 peptides (2.4%) were found in the four samples, and 212 peptides (9.1%) were exclusive of the kefir at all fermentation times, i.e. those peptides were absent in the control but found at 12, 24, and 36 h.

ACCEPTED MANUSCRIPT

3.2. Variation of protein digestion pattern over fermentation time

To gain insight into the extent of peptide release as goat milk is fermented by the kefir grains, we mapped the identified peptides on the primary sequences of the found proteins. Different patterns of sequence coverage and progression of protein digestion over fermentation time were obtained. Thus, for the caseins, a general common pattern was observed. This consisted of a variable, discontinuous coverage in the raw milk, which did not cover the whole sequence, followed by a progressive increase in primary sequence coverage throughout the fermentation time that covered approximately in a continuous way the whole mature protein at the longer fermentation times (24-36 h). In the case of α_{s1}-, α_{s2}-, and β-casein, the peptides found in the control discontinuously covered from the starting mature form of the sequences to the last amino acid in *C*-term, with gaps within the sequences that were covered in the fermented samples (Figure 2A and Supplementary Figure 1). On the contrary, for κ-casein, the sequence coverage in the control and even after 12 h fermentation was much lower, and did not reach the *C*-term region. This region, together with the larger gaps within the mature κ-casein sequence, were almost completely covered in the 24 and 36 h samples (Figure 2B).

The coverage patterns observed in the whey proteins were more variable. Thus, the highly abundant serum amyloid A protein showed a digestion pattern quite similar to those of the α- and β-casein caseins, whereas that of the GLYCAM-1 was similar to the pattern of κ-casein (Supplementary Figure 2). In a different way, β-lactoglobulin showed a very short coverage in the control (just a few amino acids at *N*-term of the mature form), which was significantly increased to reach the *C*-term, with important gaps in the middle of the sequence (Figure 2C). This pattern was just the opposite of that of osteopontin, where one of the two peptides found in the control mapped on the *C*-term (the other one mapped in the middle) and the digestion seemed to extend towards the *N*-term as the fermentation progressed (Figure 2D). However,

ACCEPTED MANUSCRIPT

other proteins like the butyrophilin subfamily 1 member A1 and lactoperoxidase, showed a very similar coverage in the control and in the fermentation samples. In these proteins, the numbers of peptides found at 12 and 36 h were less than in the control, being the highest numbers of identified peptides for the 24 h samples (Supplementary Figure 3).

3.3. Quantitative measuring of peptide changes during milk fermentation

To better understand how the progression of protein digestion was related to the appearance of their derived peptides, we quantified the changes in peptide concentrations and established models by clustering the peptides according to their concentration trends. For that, we carried out a hierarchical clustering analysis after *z*-score normalization of the peak areas in order to associate each peptide quantification to a temporal pattern. In order to reduce missingness noise, only peptides present in all the four analyzed samples including the control (n = 56) were first clustered. Additionally, a second cluster analysis was carried out for those peptides present in all the fermentation times but absent in the control sample (n = 212).

Figure 3 shows a heatmap and clustering analysis of the 56 peptides differentially abundant in the three kefir samples and also in the control. These peptides were grouped into six different clusters, according to their quantification patterns throughout the fermentation process (Supplementary Figure 4 and Supplementary Table 1). Cluster 1, composed of 20 peptides, was characterized by a slight increase at 12 and 36 h compared to the control, but a clear peak of peptide levels at 24 h. Most of those peptides (11 out of 20) corresponded to β-casein. The rest belonged to α_{s1}- and α_{s2}-casein, and one to the fibrinogen α chain. Clusters 2 (n = 10 peptides) and 4 (n = 10 peptides) were very similar in trend, with progressive increase in peptide concentration along the fermentation, but with no change in relative peptide levels up to 24 h in cluster 2, whereas in cluster 4 the increase started at that time. In both clusters, the presence of β-casein peptides was dominant, especially in cluster 4 (9 out of 10 peptides). In

ACCEPTED MANUSCRIPT

cluster 5 (*n* = 4), there was a decrease after 24 h, to reach approximately the levels of the control at 36 h. Cluster 6 (*n* = 6) showed an earlier peak at 12 h, decreasing the peptide levels thereafter. Cluster 3 (*n* = 6) had the opposite behaviour, with progressive decreases in peptide levels as fermentation progressed. However, the overall analysis shows that β-casein is highly digested over time and its derived peptides are especially abundant at 24 and 36 h.

We also compared the quantification levels of peptides found in the three fermentation times, but that were absent in non-fermented milk (Figure 4). Five different clusters were formed to group the 212 peptides included in this analysis (Supplementary Figure 5 and Supplementary Table 2). Clusters 1, 2 and 3 were those including the vast majority of the peptides (*n* = 76, *n* = 54, and *n* = 66, respectively). In these clusters, the peptide levels increased gradually from 12 to 36 h (cluster 1), or showed the highest abundance peak at 24 h (clusters 2 and 3). Again, peptides derived from β-casein were the dominant ones, but also peptides from the other caseins and from the serum amyloid A protein. Other proteins detected from less peptides than the major caseins and serum amyloid A protein, showed also peaks for their peptides preferentially at 24 h, as for GLYCAM-1 and the butyrophilin subfamily 1 member A1. Finally, clusters 4 and 5 represented decreases in peptide levels from 12 to 36 h, but these clusters contained a significantly lower number of peptides (*n* = 11, and *n* = 5, respectively).

3.4. Identification of potential bioactive peptides

Next, we sought for peptides with potential biological activity by searching in the Biopep database. In order to understand the possible limitations of our LC-MS/MS analysis when comparing the identified peptides (because of measuring peptide masses in the fixed *m/z* range) with the peptide sequences in the database, we first performed a density plot for both peptide populations. Supplementary Figure 6 shows an asymmetric beanplot that compares the peptides in the Biopep database with those that we identified experimentally. In our

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

experimental peptide set, we found peptides ranging from 5 to around 50 amino acids, but with a mode corresponding to 10 amino acids in size. Very few amino acids smaller than that size were identified, when compared to peptides of larger size. The Biopep database showed a multimodal distribution, with the main mode corresponding to 3 amino acids, and other modes of larger size. However, that mode at 2 amino acids indicates that the LC-MS/MS analysis imposes a physical limitation as the experimental technique cannot identify very short peptides.

From our experimental peptide dataset, we found 11 peptides that matched with exact primary sequences contained in the Biopep database (Table 2). Four of these peptides belonged to β-casein, 3 to κ-casein, 2 to α_{s2}-casein, 1 to α_{s1}-casein, and the other one to β-lactoglobulin. Five out of the 11 peptides were described to possess angiotensin converted enzyme (ACE) inhibitor activity; three, antibacterial activity; one was described as antioxidative; other one as antithrombotic; and the last one, as dipeptidyl peptidase IV inhibitor. Ten out of the 11 identified bioactive peptides were present in the 24-h kefir; 7 were found at 36 h; 2 were identified after 12 h fermentation. Only one bioactive peptide matching exactly with a sequence in Biopep was identified in the control, which corresponded to β-casein. This peptide was found also in all the kefir samples. Also one peptide, belonging to κ-casein, was found in all the kefir samples but not in the raw milk. This peptide had a predicted antioxidative activity. Four peptides were found only at 24 and 36 h. One peptide with ACE inhibitor activity was exclusive of the 36-h kefir, and 4 were found only after 24 h fermentation (Supplementary Figure 7). Interestingly, all the 6 peptides found in both 24 and 36 h samples were more abundant in the 36-h kefir than after 24 h fermentation, as shown by the heatmap analysis (Supplementary Figure 8).

ACCEPTED MANUSCRIPT

Finally, as the number of biopeptides found in our analysis was reduced, we screened for precursors in our experimentally identified peptides. In practice, all our peptides contained some predicted bioactive peptide, as the database contains a lot of di- and tripeptides. I.e., a peptide containing a predicted bioactive sequence (even if it is only 2 or 3 amino acid residues) was considered a biopeptide precursor. We simulated in a netplot the weighed biological activities of the biopeptide precursors, and thus, 22 different biological activities were found (Supplementary Figure 9). This analysis revealed particular enrichments of activities for each sample. Thus, after 24 h fermentation, there was an enrichment in activities like antibacterial, immunomodulating, hypotensive, anticancer, and others. The 36-h kefir was particularly enriched in activities such as ACE inhibitor, antithrombotic, antioxidative, antiamnestic, and others.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

4. Discussion

In this work, we have carried out a comprehensive analysis to profile the protein digestion and peptide occurrence in goat milk kefir throughout time, using a peptidomics approach, as a way to in-depth characterize the peptides present in a biological sample (Giacometti & Buretic-Tomljanovic, 2017). Also, we have identified peptides with potential biological activity by searching in databases. Numerous studies have explored the peptidomic profile of milk from different species including human (Dallas, Guerrero, Parker, Garay, Bhandari, Lebrilla, et al., 2014; Dallas, Smink, Robinson, Tian, Guerrero, Parker, et al., 2015; Guerrero, Dallas, Contreras, Chee, Parker, Sun, et al., 2014), as well as that of dairy products including kefir (Amorim, et al., 2019; Aspri, Leni, Galaverna, & Papademas, 2018; Dallas, et al., 2016; Ebner, Asci Arslan, Fedorova, Hoffmann, Kucukcetin, & Pischetsrieder, 2015; Piovesana, et al., 2015; Quiros, Hernandez-Ledesma, Ramos, Amigo, & Recio, 2005; Zenezini Chiozzi, Capriotti, Cavaliere, La Barbera, Piovesana, Samperi, et al., 2016). Most of the works that use proteomics/peptidomics for kefir characterization have been done on bovine milk fermented product (Amorim, et al., 2019; Dallas, et al., 2016; Ebner, Asci Arslan, Fedorova, Hoffmann, Kucukcetin, & Pischetsrieder, 2015) because of its major importance in the livestock sector worldwide. There is a lack of studies that have characterized the peptides of goat milk kefir (Quiros, Hernandez-Ledesma, Ramos, Amigo, & Recio, 2005), in spite of the importance of this ruminant species for many countries all over the World. Moreover, there is a total lack of studies of how proteins and peptides change in kefir of any species' milk throughout the fermentation process. To our knowledge, our work is the first to study at proteomic level the qualitative and qualitative changes of kefir at different fermentation times.

Considering the three fermentation times and the control analyzed in this study, we identified 2328 different unique peptides that belonged to 32 protein precursors, or 22 unique protein annotations. Of these, 15 were found in raw milk, 10 after 12 h fermentation, 20 proteins in

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

the 24 h samples and 18 in the 36-h kefir. The numbers of total unique peptides and proteins are comparable to other recent works that have used a peptidomic approach for profiling the peptides produced in kefir. Thus, Dallas et al found 2689 peptides in the four treatment groups of their experimental design in cow milk kefir samples (Dallas, et al., 2016) that corresponded to 20 protein annotations from 55 protein precursors. Very recently, Amorim et al have found in bovine milk kefir 4592 peptide spectrum matches *vs* 965 in non-fermented milk, corresponding to 28 and 18 unique proteins, respectively (Amorim, et al., 2019), although a substantial difference with our work is that they digested their peptide preparation with trypsin. However, Ebner et al identified only peptides in bovine milk kefir (considering those originated by kefir grains and starter culture), and 248 in non-fermented milk. All of them corresponded to caseins, and surprisingly they did not find any peptide belonging to whey proteins (Ebner, Asci Arslan, Fedorova, Hoffmann, Kucukcetin, & Pischetsrieder, 2015). Our results are therefore in concordance with previous studies, and discrepancies with Ebner et al could be due to differences in peptide extraction methods and analysis. Also, differences in terms of peptide and protein identification, and the extension of digestion when compared to the previous cited studies might be due to variations in kefir production: thus, in our work, after fermentation, samples were kept and processed for further analysis, whereas Dallas et al, and Amorim et al, let the kefir mature for an additional 24 h period.

Although we have not measured the proteolytic activity in kefir samples, it is evident that the huge differences in peptide and protein numbers found between the samples are due to the action of microbial proteases of kefir grains, as already described (Amorim, et al., 2019; Dallas, et al., 2016). We have neither characterized microbiologically the kefir grains, as it is also well known that is composed of a symbiosis of tens of microbial species (Rosa, Dias, Grzeskowiak, Reis, Conceicao, & Peluzio, 2017). Peptides found after 24 h fermentation were 7.7-fold of control, and the increase at 36 h compared to raw milk was 5.5-fold. At 12 h, the

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

number of identified peptides was quite similar to that of control (327 *vs* 261), probably because the kefir was not yet mature and the proteolytic activity was low at that time. The lower number of peptides, as well as proteins, found at 36 h compared to 24 h (1441 *vs* 2004) could be explained by a larger extension of digestion, which might shorten the peptides generated after 24 h and could not be detectable in the MS/MS analysis. Nevertheless, the 24 h and 36 h samples shared the highest number of peptides among all the possible intersection groups (882). Most probably this is due to the very high abundance of peptides from the major proteins, mainly caseins and the serum amyloid A protein, which deliver peptides at high concentrations and, although most of them are likely degraded as the fermentation goes on, there might be always sufficient molecules to be detected even though many of them are digested into shorter pieces. Moreover, this could also contribute to a higher coverage of the primary sequence of these proteins, compared to other proteins for which the numbers of peptides found are lesser. Also, differences in tertiary structure of proteins across species could contribute to differential sensitivity to microbial proteases digestion: bovine β-lactoglobulin was previously described to be resistant to proteolysis (Dallas, et al., 2016). However, in our study this protein was extensively digested, although in a lesser extent than caseins.

Milk is a fluid secreted by the mammary glands, with a complex composition consisting of water, fat, sugars, minerals, proteins, whole cells and other components (Roncada, Piras, Soggiu, Turk, Urbani, & Bonizzi, 2012). The protein fraction is usually divided into insoluble and soluble, i.e. the whey. Insoluble proteins are mainly the caseins, which represent up to 80% of total protein mass. Whey soluble proteins are quite variable, including many minor proteins, enzymes, hormones, and immunoglobulins. Proteomic analysis of milk has revealed the presence of hundreds of different proteins in different mammalian species after trypsin digestion of milk protein fractions (Anagnostopoulos, Katsafadou, Pierros, Kontopodis,

ACCEPTED MANUSCRIPT

Fthenakis, Arsenos, et al., 2016). Obviously, the number of different proteins is much lower when applying a peptidomic approach to identify endogenously formed peptides, for which no trypsin is used in the experimental workflow. This is the approach used in the present work, therefore the number of expected proteins identified in milk or fermented products is more limited if no trypsin is used than when the milk/fermented product is proteolytically digested prior to the LC-MS/MS analysis.

We identified a total of 22 different unique proteins, in line with most of the published works aiming at identifying endogenously released peptides, either in milk itself or by the action of microbial proteases. We found as expected all the caseins, which were identified from the vast majority of peptides found in our analysis (1693 out of 2328, i.e. 72.7% of total peptides). In addition, we found other proteins from a variable number of peptides, many of which have been also identified in cow milk kefir (Dallas, et al., 2016): α-lactalbumin, β-lactoglobulin, the serum amyloid A protein, the butyrophilin subfamily 1 member A1, GLYCAM-1, lactoperoxidase, osteopontin, perilipin, the β-1,4-galactosyltransferase I, and the parathyroid hormone-like protein.

The remaining 8 proteins were found from a low number of peptides. This low number could be explained by two facts: i) proteins are very low abundant in the original milk and/or kefir; ii) they do not release enough peptides in the milk or are resistant to microbial proteases. There is no previous evidence for the detection of these proteins in kefir, but they have been identified in milk or related biological sources.

Three cathelicidins were identified in this study, namely Bac7.5, cathelicidin 2 (Bac5) and MAP28 protein. Cathelicidins are cationic, small antimicrobial peptides produced by many vertebrates, as part of the innate immune system. They show broad antimicrobial spectrum. They are stored in neutrophil granules as inactive precursors, and released in their mature

ACCEPTED MANUSCRIPT

form when required (Kosciuczuk, Lisowski, Jarczak, Strzalkowska, Jozwik, Horbanczuk, et al., 2012). Cathelicidins have been demonstrated to have increased levels in animals with clinical mastitis (Addis, Bronzo, Puggioni, Cacciotto, Tedde, Pagnozzi, et al., 2017; Olumee-Shabon, Swain, Smith, Tall, & Boehmer, 2013), although they are also present in milk of healthy animals (Anagnostopoulos, et al., 2016; Cunsolo, Fasoli, Saletti, Muccilli, Gallina, Righetti, et al., 2015).

The fatty acid binding protein 3 has also been identified in goat milk (Cunsolo, et al., 2015; Olumee-Shabon, Swain, Smith, Tall, & Boehmer, 2013), as well as the cysteine rich secretory protein 3 (Cunsolo, et al., 2015), the fibrinogen α chain and the fatty acid synthase (Anagnostopoulos, et al., 2016), which is a key enzyme for the synthesis of milk fat in the ruminant mammary gland (Han, Gao, Yang, & Loor, 2018). Finally, the agouti signaling protein has a role in *de novo* synthesis of milk fatty acids, as a regulator of lipid metabolism in adipocytes (Knutsen, Olsen, Tafintseva, Svendsen, Kohler, Kent, et al., 2018).

Milk and derived dairy products are a source of molecules with biological activities and potential health benefits. Many of these molecules are peptides that are either present in the milk, or resulting from proteolysis of milk proteins (Hayes, Ross, Fitzgerald, & Stanton, 2007; Hayes, Stanton, Fitzgerald, & Ross, 2007). Microbial proteases contribute to the release of the vast majority of peptides with biological activity in fermented milks (Amorim, et al., 2019; Dallas, et al., 2016; Piovesana, et al., 2015). However, many bioactive sequences may be encrypted within released peptides, and the release of the definitive active form may occur after gastrointestinal digestion or further fermentation (Zenezini Chiozzi, et al., 2016). We identified 11 biopeptides in our analysis that matched exactly sequences in the Biopep database. Almost all of them (10 out of 11) belonged to casein proteins, and almost half of them (5 out of 11) were described to have ACE inhibitory activity. The highest number of predicted biopeptides was found after 24 h fermentation (10 out of 11), decreasing at 36 h

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

probably because of longer exposure to microbial proteases. Therefore, our study indicates that for goat milk kefir, fermentation at 24 h maximizes the variety of peptides with potential biological activity.

The lower number of peptides with biological activity found in our work, compared to other kefir studies (Amorim, et al., 2019; Dallas, et al., 2016), might be due to the following reasons: i) there is a predominance of very short bioactive peptides in the Biopep database, which may bias the matches; ii) most of the milk bioactive peptides in databases correspond to bovine milk, with a lesser predominance of other species; and iii) we restricted our search to exact matches with database entries. However, we cannot discard that the differences in the sequences of caseins and other proteins might cause a lack of matches between goat and cow proteins. In addition, many peptide biological activities may be hidden within peptides as precursors, which could be in the active form after additional proteolysis. Further research would be needed to simulate gastrointestinal digestion in order to study additional release of peptides from those already produced by kefir microbial proteases, and to explore their biological activities.

5. Conclusions

This is the most comprehensive proteomic analysis of goat milk kefir, and the first work in profiling the peptides released from milk proteins throughout fermentation time. Our study shows also the patterns of digestion of identified proteins, and which peptides change at the different fermentation times. Proteins and peptides identified are maximum after 24 h fermentation, and decreases at 36 h. The presence of other minor proteins and peptides could be explored in further research works by employing combinatorial peptide ligand library technologies. Moreover, it is also demonstrated the presence of peptides with biological activity and how they change along fermentation time. Many biological activities of peptides

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

are contained within other identified sequences as precursors. Further research using simulated gastrointestinal systems would be helpful to understand how biopeptide precursors of fermented beverages may become active. In addition, further research should study the biological activity of peptides appearing along time which are not included in databases. This knowledge may be used for new formulations of functional foods.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgments

LC-MS/MS and bioinformatics analyses were performed respectively at the Proteomics and Bioinformatics Units, SCAI (Central Facilities for Research Support), University of Córdoba. We are indebted to members of the AGR-164 group, headed by Prof. Jesús V. Jorrín-Novo, University of Córdoba, for lab support.

References

Addis, M. F., Bronzo, V., Puggioni, G. M. G., Cacciotto, C., Tedde, V., Pagnozzi, D., Locatelli, C., Casula, A., Curone, G., Uzzau, S., & Moroni, P. (2017). Relationship between milk cathelicidin abundance and microbiologic culture in clinical mastitis. *J Dairy Sci, 100*(4), 2944-2953.

Amorim, F. G., Coitinho, L. B., Dias, A. T., Friques, A. G. F., Monteiro, B. L., Rezende, L. C. D., Pereira, T. M. C., Campagnaro, B. P., De Pauw, E., Vasquez, E. C., & Quinton, L. (2019). Identification of new bioactive peptides from Kefir milk through proteopeptidomics: Bioprospection of antihypertensive molecules. *Food Chem, 282*, 109-119.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Anagnostopoulos, A. K., Katsafadou, A. I., Pierros, V., Kontopodis, E., Fthenakis, G. C., Arsenos, G., Karkabounas, S. C., Tzora, A., Skoufos, I., & Tsangaris, G. T. (2016). Milk of Greek sheep and goat breeds; characterization by means of proteomics. *J Proteomics, 147*, 76-84.

Aspri, M., Leni, G., Galaverna, G., & Papademas, P. (2018). Bioactive properties of fermented donkey milk, before and after in vitro simulated gastrointestinal digestion. *Food Chem, 268*, 476-484.

Clark, S., & Mora Garcia, M. B. (2017). A 100-Year Review: Advances in goat milk research. *J Dairy Sci, 100*(12), 10026-10044.

Contreras, M. M., Carrón, R., Montero, M. J., Ramos, M., & Recio, I. (2009). Novel casein-derived peptides with antihypertensive activity. *Int Dairy J, 19*, 566-573.

Cunsolo, V., Fasoli, E., Saletti, R., Muccilli, V., Gallina, S., Righetti, P. G., & Foti, S. (2015). Zeus, Aesculapius, Amalthea and the proteome of goat milk. *J Proteomics, 128*, 69-82.

Dallas, D. C., Citerne, F., Tian, T., Silva, V. L., Kalanetra, K. M., Frese, S. A., Robinson, R. C., Mills, D. A., & Barile, D. (2016). Peptidomic analysis reveals proteolytic activity of kefir microorganisms on bovine milk proteins. *Food Chem, 197*(Pt A), 273-284.

Dallas, D. C., Guerrero, A., Parker, E. A., Garay, L. A., Bhandari, A., Lebrilla, C. B., Barile, D., & German, J. B. (2014). Peptidomic profile of milk of Holstein cows at peak lactation. *J Agric Food Chem, 62*(1), 58-65.

Dallas, D. C., Smink, C. J., Robinson, R. C., Tian, T., Guerrero, A., Parker, E. A., Smilowitz, J. T., Hettinga, K. A., Underwood, M. A., Lebrilla, C. B., German, J. B., & Barile, D. (2015). Endogenous human milk peptide release is greater after preterm birth than term birth. *J Nutr, 145*(3), 425-433.

ACCEPTED MANUSCRIPT

de Oliveira Leite, A. M., Miguel, M. A., Peixoto, R. S., Rosado, A. S., Silva, J. T., & Paschoalin, V. M. (2013). Microbiological, technological and therapeutic properties of kefir: a natural probiotic beverage. *Braz J Microbiol, 44*(2), 341-349.

Deutsch, E. W., Csordas, A., Sun, Z., Jarnuczak, A., Perez-Riverol, Y., Ternent, T., Campbell, D. S., Bernal-Llinares, M., Okuda, S., Kawano, S., Moritz, R. L., Carver, J. J., Wang, M., Ishihama, Y., Bandeira, N., Hermjakob, H., & Vizcaino, J. A. (2017). The ProteomeXchange consortium in 2017: supporting the cultural change in proteomics public data deposition. *Nucleic Acids Res, 45*(D1), D1100-D1106.

Ebner, J., Asci Arslan, A., Fedorova, M., Hoffmann, R., Kucukcetin, A., & Pischetsrieder, M. (2015). Peptide profiling of bovine kefir reveals 236 unique peptides released from caseins during its production by starter culture or kefir grains. *J Proteomics, 117*, 41-57.

Escareño, L., Salinas-Gonzalez, H., Wurzinger, M., Iniguez, L., Solkner, J., & Meza-Herrera, C. (2013). Dairy goat production systems: status quo, perspectives and challenges. *Trop Anim Health Prod, 45*(1), 17-34.

FAOSTAT. (2017). http://faostat.fao.org/site/569/DesktopDefault.aspx?PageID=569 - ancor.

Giacometti, J., & Buretic-Tomljanovic, A. (2017). Peptidomics as a tool for characterizing bioactive milk peptides. *Food Chem, 230*, 91-98.

Guerrero, A., Dallas, D. C., Contreras, S., Chee, S., Parker, E. A., Sun, X., Dimapasoc, L., Barile, D., German, J. B., & Lebrilla, C. B. (2014). Mechanistic peptidomics: factors that dictate specificity in the formation of endogenous peptides in human milk. *Mol Cell Proteomics, 13*(12), 3343-3351.

Han, L. Q., Gao, T. Y., Yang, G. Y., & Loor, J. J. (2018). Overexpression of SREBF chaperone (SCAP) enhances nuclear SREBP1 translocation to upregulate fatty acid synthase

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

(FASN) gene expression in bovine mammary epithelial cells. *J Dairy Sci, 101*(7), 6523-6531.

Hayes, M., Ross, R. P., Fitzgerald, G. F., Hill, C., & Stanton, C. (2006). Casein-derived antimicrobial peptides generated by Lactobacillus acidophilus DPC6026. *Appl Environ Microbiol, 72*(3), 2260-2264.

Hayes, M., Ross, R. P., Fitzgerald, G. F., & Stanton, C. (2007). Putting microbes to work: dairy fermentation, cell factories and bioactive peptides. Part I: overview. *Biotechnol J, 2*(4), 426-434.

Hayes, M., Stanton, C., Fitzgerald, G. F., & Ross, R. P. (2007). Putting microbes to work: dairy fermentation, cell factories and bioactive peptides. Part II: bioactive peptide functions. *Biotechnol J, 2*(4), 435-449.

Klippel, B. F., Duemke, L. B., Leal, M. A., Friques, A. G., Dantas, E. M., Dalvi, R. F., Gava, A. L., Pereira, T. M., Andrade, T. U., Meyrelles, S. S., Campagnaro, B. P., & Vasquez, E. C. (2016). Effects of Kefir on the Cardiac Autonomic Tones and Baroreflex Sensitivity in Spontaneously Hypertensive Rats. *Front Physiol, 7*, 211.

Knutsen, T. M., Olsen, H. G., Tafintseva, V., Svendsen, M., Kohler, A., Kent, M. P., & Lien, S. (2018). Unravelling genetic variation underlying de novo-synthesis of bovine milk fatty acids. *Sci Rep, 8*(1), 2179.

Korhonen, H., & Pihlanto, A. (2007). Technological options for the production of health-promoting proteins and peptides derived from milk and colostrum. *Curr Pharm Des, 13*(8), 829-843.

Kosciuczuk, E. M., Lisowski, P., Jarczak, J., Strzalkowska, N., Jozwik, A., Horbanczuk, J., Krzyzewski, J., Zwierzchowski, L., & Bagnicka, E. (2012). Cathelicidins: family of antimicrobial peptides. A review. *Mol Biol Rep, 39*(12), 10957-10970.

ACCEPTED MANUSCRIPT

Lopez-Exposito, I., Gomez-Ruiz, J. A., Amigo, L., & Recio, I. (2006). Identification of antibacterial peptides from ovine alpha s2-casein. *Int Dairy J, 16*, 1072-1080.

Olumee-Shabon, Z., Swain, T., Smith, E. A., Tall, E., & Boehmer, J. L. (2013). Proteomic analysis of differentially expressed proteins in caprine milk during experimentally induced endotoxin mastitis. *J Dairy Sci, 96*(5), 2903-2912.

Pellegrini, A., Dettling, C., Thomas, U., & Hunziker, P. (2001). Isolation and characterization of four bactericidal domains in the bovine beta-lactoglobulin. *Biochim Biophys Acta, 1526*(2), 131-140.

Piovesana, S., Capriotti, A. L., Cavaliere, C., La Barbera, G., Samperi, R., Zenezini Chiozzi, R., & Lagana, A. (2015). Peptidome characterization and bioactivity analysis of donkey milk. *J Proteomics, 119*, 21-29.

Qian, Z. Y., Jolles, P., Migliore-Samour, D., & Fiat, A. M. (1995). Isolation and characterization of sheep lactoferrin, an inhibitor of platelet aggregation and comparison with human lactoferrin. *Biochim Biophys Acta, 1243*(1), 25-32.

Quiros, A., del Mar Contreras, M., Ramos, M., Amigo, L., & Recio, I. (2009). Stability to gastrointestinal enzymes and structure-activity relationship of beta-casein-peptides with antihypertensive properties. *Peptides, 30*(10), 1848-1853.

Quiros, A., Hernandez-Ledesma, B., Ramos, M., Amigo, L., & Recio, I. (2005). Angiotensin-converting enzyme inhibitory activity of peptides derived from caprine kefir. *J Dairy Sci, 88*(10), 3480-3487.

Rodriguez-Ortega, M. J. (2018). "Shaving" Live Bacterial Cells with Proteases for Proteomic Analysis of Surface Proteins. *Methods Mol Biol, 1722*, 21-29.

Roncada, P., Piras, C., Soggiu, A., Turk, R., Urbani, A., & Bonizzi, L. (2012). Farm animal milk proteomics. *J Proteomics, 75*(14), 4259-4274.

ACCEPTED MANUSCRIPT

Rosa, D. D., Dias, M. M. S., Grzeskowiak, L. M., Reis, S. A., Conceicao, L. L., & Peluzio, M. (2017). Milk kefir: nutritional, microbiological and health benefits. *Nutr Res Rev, 30*(1), 82-96.

Selvaggi, M., Laudadio, V., Dario, C., & Tufarelli, V. (2014). Major proteins in goat milk: an updated overview on genetic variability. *Mol Biol Rep, 41*(2), 1035-1048.

Villegas, J. M., Picariello, G., Mamone, G., Espeche Turbay, M. B., Savoy de Giori, G., & Hebert, E. M. (2014). Milk-derived angiotensin-I-converting enzyme-inhibitory peptides generated by Lactobacillus delbrueckii subsp. lactis CRL 581. *Peptidomics, 1*, 22-29.

Vizcaino, J. A., Csordas, A., del-Toro, N., Dianes, J. A., Griss, J., Lavidas, I., Mayer, G., Perez-Riverol, Y., Reisinger, F., Ternent, T., Xu, Q. W., Wang, R., & Hermjakob, H. (2016). 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res, 44*(D1), D447-456.

Zenezini Chiozzi, R., Capriotti, A. L., Cavaliere, C., La Barbera, G., Piovesana, S., Samperi, R., & Lagana, A. (2016). Purification and identification of endogenous antioxidant and ACE-inhibitory peptides from donkey milk by multidimensional liquid chromatography and nanoHPLC-high resolution mass spectrometry. *Anal Bioanal Chem, 408*(20), 5657-5666.

Zhang, Y., Chen, R., Ma, H., & Chen, S. (2015). Isolation and Identification of Dipeptidyl Peptidase IV-Inhibitory Peptides from Trypsin/Chymotrypsin-Treated Goat Milk Casein Hydrolysates by 2D-TLC and LC-MS/MS. *J Agric Food Chem, 63*(40), 8819-8828.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Figure captions

Figure 1. Intersection diagram of peptides identified in the goat milk kefir samples at the three fermentation times (12, 24, and 36 h), and the control (0 h non-fermented raw milk). Numbers on the bars represent the peptides belonging to each intersection group.

Figure 2. Mapping of identified peptides for the four samples (0 h non-fermented raw milk, and fermentation times at 12, 24, and 36 h) on the sequences of A) β-casein; B) κ-casein; C) β-lactoglobulin; and D) osteopontin.

Figure 3. Heatmap and hierarchical clustering analysis of the 56 peptides identified in common in the four samples analyzed (0 h non-fermented raw milk, and fermentation times at 12, 24, and 36 h).

Figure 4. Heatmap and hierarchical clustering analysis of the 212 peptides identified in common in the three kefir samples (12, 24, and 36 h fermentation times), but not in the control raw milk.

Legends to supplemental material

Supplementary Figure 1. Mapping of identified peptides for the four samples (non-fermented raw milk, and fermentation times at 12, 24, and 36 h) on the sequences of A) α_{s1}-casein; and B) α_{s2}-casein.

Supplementary Figure 2. Mapping of identified peptides for the four samples (non-fermented raw milk, and fermentation times at 12, 24, and 36 h) on the sequences of A) serum amyloid A protein; and B) glycosylation-dependent cell adhesion molecule 1.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Supplementary Figure 3. Mapping of identified peptides for the four samples (non-fermented raw milk, and fermentation times at 12, 24, and 36 h) on the sequences of A) butyrophilin subfamily 1 member A1; and B) lactoperoxidase.

Supplementary Figure 4. Clusters of the 56 peptides identified in common in the four samples analyzed (non-fermented raw milk, and fermentation times at 12, 24, and 36 h).

Supplementary Figure 5. Clusters of the 212 peptides identified in common in the three kefir samples (12, 24, and 36 h fermentation times), but not in the control raw milk.

Supplementary Figure 6. Beanplot of the Biopep database peptide population *vs* the peptide population experimentally identified in this work.

Supplementary Figure 7. Intersection diagram of bioactive peptides identified in the goat milk kefir samples at the three fermentation times (12, 24, and 36 h), and the control (non-fermented raw milk).

Supplementary Figure 8. Heatmap and hierarchical clustering analysis of the 8 bioactive peptides identified in the four samples analyzed (non-fermented raw milk, and fermentation times at 12, 24, and 36 h).

Supplementary Figure 9. Netplot analysis of the bioactivity of precursor sequences identified in the goat milk kefir samples at the three fermentation times (12, 24, and 36 h), and the control (non-fermented raw milk).

Supplementary Dataset 1. List of peptides identified in all the biological replicates of the non-fermented raw milk and in kefir at the three fermentation times (12, 24, and 36 h).

Supplementary Table 1. z-score values of the 56 peptides identified in common in the four samples analyzed (0 h-control raw milk, and 12, 24 and 36 h kefir) for the clustering analysis.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Supplementary Table 2. z-score values of the 212 peptides identified in common in the three kefir samples analyzed (12, 24 and 36 h) for the clustering analysis.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Table 1. identified proteins and peptides in the the three kefir samples (12, 24, and 36 h fermentation time) and the control (non-fermented raw milk)

Protein annotation	Uniprot entries	# Uniprot entries	# Peptides (control)	# Peptides (12 h kefir)	# Peptide (24 h kef
Agouti signaling protein	A9JPS5	1	0	0	1
α-lactalbumin	A5JSS8	1	0	0	8
α_{S1}-casein	A0A0P0E041,A0A0P0EL46, A0A0P0EUH2,P18626,Q8MIH4	5	47	45	248
α_{S2}-casein	P33049	1	45	38	275
Bac7.5 protein	Q9XSQ9	1	0	0	2
β-1,4-galactosyltransferase I	E9NRZ3	1	1	0	2
β-casein	P33048,Q95L76	2	69	151	762
β-lactoglobulin	P02756	1	2	0	30
Butyrophilin subfamily 1 member A1	A3EY52	1	27	11	49
Cathelicidin 2	P82018	1	0	0	2
Cysteine rich secretory protein 3	D6PX62	1	0	0	0
Fatty acid binding protein 3	Q6S4N9	1	0	0	1
Fatty acid synthase	Q06B57	1	1	0	0
Fibrinogen alpha chain	P68215	1	2	1	2
Glycosylation dependent cell adhesion molecule 1	A5JSS7	1	7	9	58
κ-casein	A0SWR2,Q5VJR8,Q7YRV1,Q7YRX4,Q7YRX5	5	7	21	187
Lactoperoxidase	A3F9D6	1	12	5	27
MAP28 protein	Q9XSQ8	1	0	0	1
Osteopontin	U5Y6U2	1	2	6	18
Parathyroid hormone like protein	E2EAI4	1	1	0	2
Perilipin	G3EHG6	1	1	0	2
Serum amyloid A protein	A5JSR9,A5JST2	2	37	40	327
TOTAL		32	261	327	2004

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Table 2. Identified peptides that matched exactly in sequence and length to entries in the Biopep database.

Sequence	Protein accession and position	Protein annotation	Activity	Sample[a]	Reference
AASDISLLDAQSAPLR	P02756 [43-58]	β-lactoglobulin	antibacterial	24	Pellegrini et al, 2001
ARHPHPHLSFM	Q7YRX5 [87-97]; Q5VJR8 [87-97]; A0SWR2 [87-97]; Q7YRX4 [87-97]; Q7YRV1 [87-97]	κ-casein	antioxidative	12, 24, 36	Korhonen & Pihlanto, 2007
INNQFLPYPY	Q7YRX5 [42-51]; Q5VJR8 [42-51]; Q7YRX4 [42-51]; Q7YRV1 [42-51]	κ-casein	dipeptidyl peptidase IV inhibitor	24, 36	Zhang et al, 2015
LGPVRGPFP	Q95L76 [211-219]; P33048 [211-219]	β-casein	ACE inhibitor	24	Villegas et al, 2014
LTLTDVE	Q95L76 [140-146]; P33048 [140-146]	β-casein	ACE inhibitor	24	Hayes, Stanton, Fitzgerald & Ross, 2007
QEPVLGPVRGPFP	Q95L76 [207-219]; P33048 [207-219]	β-casein	ACE inhibitor	0, 12, 24, 36	Villegas et al, 2014
TAQVTSTEV	Q7YRX5 [154-162]; Q5VJR8 [154-162]; Q7YRX4	κ-casein	antithrombotic	24, 36	Qian et al, 1995

ACCEPTED MANUSCRIPT

	[154-162]; Q7YRV1 [154-162]				
VDQHQKAMKPWTQPKTNAIPYVRYL	P33049 [199-223]	α_{S2}-casein	antibacterial	24	López-Expósito et al, 2006
VLNENLLR	Q8MIH4 [30-37]; A0A0P0EUH2 [29-36]; A0A0P0E041 [29-36]	α_{S1}-casein	antibacterial	24, 36	Hayes et al, 2006
YQKFPQY	P33049 [105-111]	α_{S2}-casein	ACE inhibitor	24, 36	Contreras et al, 2009
LTQTPVVVPPF	Q95L76 [92-102]; P33048 [92-102]	β-casein	ACE inhibitor	36	Quirós et al, 2009

[a]0: control, non-fermented raw milk; 12: 12-h fermented kefir; 24: 24-h fermented kefir; 36: 36-h fermented kefir.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Highlights

- A comprehensive peptidomic analysis of goat milk kefir at different fermentation times.
- Establishment of protein digestion patterns throughout the milk fermentation.
- Quantification of occurring peptides along the fermentation process.
- Identification of peptides with potential biological activity in goat milk kefir.

ACCEPTED MANUSCRIPT