JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

Article

Predicting the important enzyme players in human breast milk digestion

Nora Khaldi, Vaishnavi Vijayakumar, David C. Dallas, Andrés Guerrero, Saumya Wickramasinghe, Jennifer T. Smilowitz, Juan F. Medrano, Carlito B. Lebrilla, Denis C Shields, and J. Bruce German

J. Agric. Food Chem., **Just Accepted Manuscript** • DOI: 10.1021/jf405601e • Publication Date (Web): 12 Mar 2014

Journal of Agricultural and Food Chemistry is published by the American Chemical Society. 1155 Sixteenth Street N.W., Washington, DC 20036

Title. Predicting the important enzyme players in human breast milk digestion

Nora Khaldi[a,b,c,1], Vaishnavi Vijayakumar[a,b], David C. Dallas[c,d], Andrés Guerrero[e], Saumya Wickramasinghe[f,g], Jennifer T. Smilowitz[c,d], Juan F. Medrano[g], Carlito B. Lebrilla[e], Denis C. Shields[a,b], and J. Bruce German[c,d]

[a]Conway Institute of Biomolecular and Biomedical Research, School of Medicine and Medical Sciences. University College Dublin, Dublin, Republic of Ireland

[b]Complex and Adaptive Systems Laboratory, University College Dublin, Dublin, Republic of Ireland

[c]Department of Food Science and Technology, University of California, Davis, California 95616, USA.

[d]Foods for Health Institute, University of California, Davis, California 95616, USA.

[e]Department of Chemistry University of California, Davis Califorina 95616, USA.

[f]Department of Basic Veterinary Sciences, University of Peradeniya, Sri Lanka

[g]Department of Animal Science, University of California, Davis, California 95616, USA.

[1]To whom correspondence should be addressed: Nora Khaldi, nora.khaldi@ucd.ie, phone: 00353 -1-7165335, UCD Complex and Adaptive Systems Laboratory, University College Dublin, Dublin 4, Ireland.

ABSTRACT

Human milk is known to contain several proteases, but little is known about whether these enzymes are active, which proteins they cleave and their relative contribution to milk protein digestion *in vivo*. We analyzed the mass spectrometry-identified protein fragments found in pooled human milk by comparing their cleavage sites with the enzyme specificity patterns of an array of enzymes. The results indicate that several enzymes are actively taking part in the digestion of human milk proteins within the mammary gland, including plasmin and/or trypsin, elastase, cathepsin D, pepsin, chymotrypsin, a glutamyl endopeptidase-like enzyme and proline endopeptidase. Two proteins were most affected by enzyme hydrolysis: β-casein and polymeric immunoglobulin receptor. In contrast, other highly abundant milk proteins such as α-lactalbumin and lactoferrin appear to have undergone no proteolytic cleavage. We also show that a peptide sequence containing a known anti-microbial peptide is released in breast milk by elastase and cathepsin D.

KEYWORDS: hydrolysate, human milk digestion, milk, nutrition, proteolytic enzymes, bioactive peptide.

INTRODUCTION

Milk is a live secretion that contains numerous complex biomolecules such as proteins, oligosaccharides and lipids. In addition to these components, previous studies have shown that milk also contains active and inactive forms of hydrolytic enzymes capable of acting upon these biopolymers, including proteolytic enzymes. These independent studies have shown that milk contains plasmin (*1, 2*); immunoreactive anionic trypsin, most likely present in complex with IgA (*3, 4*); and cathepsin D (*5*).

Although it has been established that proteolytic enzymes are present in milk, little is known about whether these enzymes are active on milk proteins. In addition, if the enzymes were to be active, their contribution to milk protein hydrolysis, and their relative contribution compared to one another are unknown.

Many of the studies of milk proteolytic enzymes were carried out on bovine milk, mainly for cheese- or mastitis-related questions; very few were carried out on human milk (*6-9*). Most milk proteolytic enzyme investigations have been carried out separately for each enzyme and do not determine the specific effects of each enzyme on milk proteins. This lack of research means that the impact the enzymes have on human milk proteins remains largely unknown. The biological value of these proteolytic enzymes in milk remains ambiguous. Some reports suggest that milk enzymes may support infant growth and nutrition (*10*). Some studies suggest that these enzymes may release bioactive peptides that are beneficial for the infant's development (see reviews (*11, 12*)). But, so far, none of these peptide actions and *in vivo* catalytic releases have been demonstrated.

Understanding milk composition, and its diverse catalytic activities, notably the proteolytic enzymes, will shed light on mammalian development, evolution, and

how to protect neonates. A new generation of analytical and computational tools has made it possible to investigate milk's biopolymers in greater breadth and detail. We previously published a novel mass spectrometry (MS)-based peptide search platform that we used in an iterative searching strategy to identify peptides in minor abundances in human milk (*13*). We discovered many peptides, proving that milk proteins can be digested prior entering the infant's digestive system. Many of these peptides overlapped with known antimicrobial peptides (*13*). In this study, we assess the proteolytic activity in human breast milk. We used computational tools to analyze the peptide fragments given their milk protein sequence context. Using these tools, we predicted the proteolytic enzymes that are active in human milk, we examined the relative contribution of each enzyme, and we identified the proteins that have been the most cleaved by the predicted enzymes. mRNA expression was also measured in human milk to determine whether these enzymes are expressed in the mammary gland or whether they enter milk from other sources such as blood.

MATERIALS AND METHODS

Breast milk collection: Mature breast milk was collected from two healthy mothers who delivered at term (IRB number 216198). The milk was kept in storage up to 1 month. There is no impact on milk proteins and enzymes with storage. In other words, no enzyme activity occurs when samples are frozen. The stage of lactation of the first mother was 3 months and the second was 4 months. The mothers providing this milk were instructed to cleanse each breast with water and then use an electronic pump to collect a pool of milk from both breasts. The samples were then stored in the subjects' freezers and delivered to the laboratory on ice where they were stored at -80C.

Sample preparation: Peptides were extracted from human milk according to the method of Dallas *et al.* (*13*). Briefly, 200 µL of each milk sample were thawed on ice and combined. To remove milk fat globules, the milk was centrifuged at 3,000 x g for 10 min and the skim infranate was extracted. Centrifugation was repeated on the skim infranate to remove any remaining visible lipid layer. Proteins were then precipitated with addition of 400 µL of 200 g/L trichloroacetic acid. Samples were vortexed briefly, then centrifuged at 3,000 x g for 10 min and the peptide-containing supernatant was collected, leaving the precipitated protein. This precipitation was repeated for a total of three times. Trichloroacetic acid, salts, oligosaccharides and lactose were then removed from the peptides by C18 reverse-phase preparative chromatography. Contaminants were eluted with water and peptides were then eluted with 80% acetonitrile (ACN), 0.1 % trifluoroacetic acid (v/v). The peptide solution was then dried down in a vacuum centrifuge at 37ºC. After drying, the sample was rehydrated in 40 µL of nanopure water for MS analysis.

MS analysis of human breast milk peptides: Peptides were analyzed via nano-liquid chromatography chip quadrupole time-of-flight tandem MS (Agilent, Santa Clara, CA). Two μL of sample were injected for each run onto the C18 reverse-phase nano chip. The nanopump flow was .3 μL/min and the capillary pump flow rate was 3 μL/min. Peptides were eluted with the following gradient of solvent A (3% ACN, 0.1% formic acid (FA) (v/v)) and solvent B (90% ACN, 0.1% FA (v/v)): 0–8% B from 0–5 min, 8–26.5% B from 5–24 min, 26.5–100% B from 24–48 min, followed by 100% B for 2 min and 100% A for 10 min (to re-equilibrate the column). The instrument was run in positive ionization mode. Data collection thresholds were set at 200 ion counts or 0.01% relative intensity for MS spectra and 5 ion counts or 0.01% relative intensity for MS/MS. Data were collected in centroid mode. The drying gas was 350 °C and flow rate was 3 L/min. The required chip voltage for consistent spray varied from 1850 to 1920 V. Automated precursor selection based on abundance was employed to select peaks for tandem fragmentation with an exclusion list consisting of all peptides identified in previous analyses in this study. The acquisition rate employed was 3 spectra/s for both MS and MS/MS modes. The isolation width for tandem analysis was 1.3 m/z. The collision energy was set by the formula (Slope)*(m/z)/100+Offset, with slope = 3.6 and offset = -4.8. Five tandem spectra were collected after each MS spectrum, with active exclusion after 5 MS/MS for 0.15 min. Precursor ions were only selected if they had at least 1000 ion counts or 0.01% of the relative intensity of the spectra. Mass calibration was performed during data acquisition based on an infused calibrant ion with a mass of 922.009789 Da. Agilent Mass Hunter Qualitative Analysis Software (Santa Clara, CA) was used to analyze the data obtained. Molecules identified in the spectral analysis were grouped into compounds by the Find by Molecular Feature algorithm, which groups together

132 molecules across charge state and charge carrier. All tandem-MS from each data file
133 were exported as Mascot Generic Files (.mgf) with a peptide isotope model and a
134 maximum charge state of +9.
135 Peptide identification was accomplished using both the MS-GFDB (via a command-
136 line interface) and X!Tandem (using the downloadable graphical user interface). The
137 human milk library used in both searches was constructed based on a query to the
138 Uniprot database. The query returned only proteins from *Homo sapiens* and at least
139 one of the following: “tissue specificity” keyword “milk” or “mammary”, “tissue”
140 keyword “milk” or “mammary” or gene ontology “lactation”. This query returned a
141 list of 1,472 proteins. These were exported to FASTA file format. For MS-GFDB,
142 peptides were accepted if p-values were less than or equal to 0.01 corresponding to
143 confidence levels of 99%. No p-values exist in X!Tandem, so a closely related
144 statistic, e-value, was used for the X!Tandem search. The e-value thresholds selected
145 were again 0.01, reflecting 99% confidence. In both programs, masses were allowed
146 20 ppm error. No complete (required) modifications were included but up to four
147 potential modifications were allowed on each peptide. Potential modifications
148 allowed were phosphorylation of serine, threonine or tyrosine and oxidation of
149 methionine. A non-specific cleavage ([X]|[X]) (where ‘X’ is any amino acid) was
150 used to search against the protein sequences. For MS-GFDB, the fragmentation
151 method selected in the search was collision-induced dissociation and the instrument
152 selected was time-of-flight. For X!Tandem, there was no option for fragmentation
153 type and instrument selection. Because the instrument did not always select the
154 monoisotopic ion for tandem fragmentation, isotope errors were allowed (allowing up
155 to one C13). No model refinement was employed in X!Tandem.
156

Enzyme prediction: The web-based software EnzymePredictor (*14*) was employed to evaluate and predict which enzymes most likely contributed to cleavage of human breast milk proteins (http://bioware.ucd.ie/~enzpred/Enzpred.php). Enzymes were classified based on their total number of performed cleavages, and they were evaluated based on their odds ratio (OR; See Table 1), which is an indicator of their degree of participation in the hydrolysis or the proteins. The OR values indicate that certain enzymes are over-represented, and others under-represented. The following information were also collected from EnzymePredictor: number of times an enzyme could have cleaved within the current peptides, total number of proteins cleaved by each enzyme, total number of cleavages performed by an enzyme on the *C*- and *N*-terminus.

Expression profiling of human milk

Fresh milk sample collection: Fresh milk samples were obtained from three healthy females on days 4, 15, 30 and 60 postpartum who gave birth to a term infant (> 37 weeks of gestation). In the early morning period, the donor manually pumped one breast until emptied into a collection bag, and immediately delivered on cold-packs to the lab for processing. The Institutional Review Board of University of California, Davis, approved the project.

RNA extraction for gene expression studies: Somatic cells were pelleted by adding 50 µL of 0.5 M ethylenediaminetetraacetic acid to 20 mL of fresh milk and centrifuged at 2,000 x g at 4˚C for 10 min (*15*). The pellet of cells was washed with 10 mL of phosphate-buffered saline at pH 7.2 and 10 µL of 0.5 M

ethylenediaminetetraacetic acid (final concentration 0.5 mM) and filtered through sterile cheesecloth to remove any debris. The cells were then centrifuged again at 2,000 x g at 4˚C for 10 min. The supernatant was decanted and RNA was extracted from the milk somatic cell pellet using Trizol (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions. RNA was quantified by an ND-1000 spectrophotometer (Fisher Thermo, Wilmington, MA) and the quality and integrity was assessed by the spectrophotometer 260/280 ratio, gel electrophoresis and by capillary electrophoresis with an Experion Bio-analyzer (Bio-Rad, Hercules, CA).

RNA sequencing and data analysis: Gene expression analysis was conducted on fresh milk samples collected on days 4, 15, 30 and 60 postpartum by RNA sequencing (RNA-Seq). Messenger RNA was isolated and purified using an RNA-Seq sample preparation kit (Illumina, San Diego, CA). The fragments were purified and sequenced at the UC Davis Genome Center DNA Technologies Core Facility using the Illumina Genome Analyzer (GAII) and Illumina HiSeq 2000. Sequence reads were assembled and analyzed in RNA-Seq. Expression analysis was performed with the CLC Genomics Workbench 6.0 (CLC Bio, Aarhus, Denmark). Human Genome, GRCh37.69 (ftp://ftp.ensembl.org/pub/release-69/genbank/homo_sapiens/) was utilized as the reference genome for the assembly. Data were normalized by calculating the 'reads per kilobase per million mapped reads' (RPKM) (*16*) for each gene and annotated with ENSEMBL human genome assembly (55,203 unique genes).

RESULTS AND DISCUSSION

This project aimed to determine the enzymes responsible for the cleavage patterns identified in milk; examine each enzyme's contribution to the hydrolyses of human milk proteins; and clarify whether these enzymes are expressed in milk or enter milk from other sources.

A novel MS-based peptide search platform using an iterative searching strategy detected a variety of peptides in minor abundances in pooled human breast milk samples (*13*). The peptides extracted and identified for the present study were analyzed using EnzymePredictor (*14*). The results of this analysis are presented in Table 1. The enzymes are ordered based on their total number of cleavages, from those that have cleaved extensively to those that are predicted to have cleaved one residue. As discussed in the interpretation of EnzymePredictor (*14*), the enzymes with a combined high number of total cleavages performed and a high odds ratio are the enzymes that are most likely active in the milk.

Expression profiles from samples of human breast milk were established as the means to investigate if the predicted enzymes originate from the mammary epithelial cells or they enter milk from other sources, such as blood. The samples that were obtained for analysis for peptide detection via MS were pooled from mature milk samples from two mothers, both 3-months postpartum, while that of expression data was of 2-month postpartum milk.

1. Major cleavage of proteins in human milk is at trypsin and plasmin target sites.

The natural cleavages of milk proteins showed a strong enrichment for cleavage after K or R, both consistent with plasmin and trypsin cleavage (Table 1,4).

The presence of plasmin—an enzyme that plays a major role in the proteolytic

229 breakdown of blood clots—has been previously reported in milk (*1*), but its
230 contribution to human milk proteolytic hydrolysis is unknown. This study found that
231 plasmin potentially cleaved 320 peptides (Fig. 1, Table 1) derived from 15 milk
232 proteins (Table 2). Anionic trypsin's presence in human milk was initially identified
233 by Monti *et al.* and Borulf *et al.* (*3, 4*). Borulf *et al.* showed that trypsin's presence in
234 milk is likely in complex with IgA. Whether this enzyme contributes to the hydrolysis
235 of human milk proteins remains unknown.

236 The two proteins most affected by potential plasmin or trypsin digestion were
237 polymeric immunoglobulin receptor (32 cleavages of peptides) and β-casein (89).
238 Other hydrolyzed proteins included α_{S1}-casein (11), osteopontin (18), butyrophilin
239 subfamily 1 member A1 (15), and κ-casein (7). These analyses found that plasmin
240 cleaves an androgen receptor that trypsin failed to cleave (Q9UN21; Table 2-3).
241 Butyrophilin-subfamily 1 member A1 is part of the milk fat globule membrane. Thus,
242 the low number of peptides detected from this protein might be due to the
243 centrifugation for the removal of the milk fat globules. Trypsin can typically not
244 cleave if the K or R is followed by a P, which makes the androgen receptor better fit
245 the plasmin pattern (Table 4). This result would suggest that the cleavage tends to
246 follow the specificity of plasmin more consistently than that of trypsin. However,
247 there are too few cleavages representing K or R followed by P (only one in our case)
248 to disambiguate between plasmin and trypsin activities in the milk.

249 **2. Enzymes cleaving hydrophobic target sites: elastase, cathepsin D, pepsin and**
250 **chymotrypsin.**

251 While the active form of cathepsin D is found in bovine milk (*5*), the major form of
252 this enzyme in milk is the inactive zymogen, procathepsin D (*17*). In this study,
253 cathepsin D was found to be active in human milk (Table 1), cleaving 130 times.

Cathepsin D hydrolyzed the highest number of milk proteins (24 milk proteins) compared to the other enzymes present in milk (Table 1,5; Figure 1). Although the total number of times cathepsin D performed a cleavage is high (130 total cleavages), most of these are located in two proteins: β-casein with 78 cleavages and polymeric immunoglobulin receptor with 21 cleavages. Interestingly, both these proteins have also been the most digested ones with all 5 top predicted enzymes.

A high number of potential cleavage sites was predicted for cathepsin D (846; Table 1), which were not observed in the peptide results, with an Odds Ratio below one. This failure to produce predicted peptides could be due to the positioning of a sub-population of these residues in regions difficult for the enzyme to access due to structural constraints. This structural inhibition is more likely for cathepsin D sites, since many potential cleavage sites are strongly hydrophobic, and hydrophobic regions are usually buried within the structured regions of proteins.

Elastase is another enzyme predicted to be active in human milk. Elastase is known to be a major proteinase in polymorphonuclear neutrophils (PMNs), which are phagocytic cells that destroy infectious agents in humans (*18*). Elastase activity has been detected in PMNs recovered from milk during experimentally induced mastitis (*19*). There is substantial overlap between the cleavage sites of elastase and cathepsin D, and again there are a large number of sites (563) interior to the peptides that were not cleaved. Elastase potentially performed 90 total cleavages of milk peptides (Table 6) over 13 proteins. The proteins that elastase cleaved the most are β-casein (41 cleavages of unique peptides) and polymeric immunoglobulin receptor (25 cleavages of unique peptides). No traces of elastase activity were observed on κ-casein (Table 6), and only one cleavage for cathepsin D (Table 5). This reflects the similar, but not identical, cleavage pattern between elastase and cathepsin D. To resolve the

specificity, the number of sites that were cleaved in common between both were determined (60 cleavages of unique peptides) or specific to only one (69 for cathepsin D of unique peptides, and 28 for elastase of unique peptides). This pattern suggests a role for both enzymes in human milk digestion prior to infant consumption.

Two other enzymes with hydrophobic targets, pepsin and chymotrypsin, are consistent with a number of digestion sites additional to those digestible by cathepsin D and elastase. Only 21 of the 75 total peptide terminal cleavage sites of unique peptides that are predicted to be performed by chymotrypsin are common cleavages with elastase and cathepsin D. For pepsin, only 38 of the 72 cleavages sites predicted from the unique peptides are shared with elastase and cathepsin D. Furthermore, only 16 peptide terminal cleavage sites are shared between both pepsin and chymotryspin. These findings indicate the first support that pepsin- and chymotrypsin-like activities are present in human breast milk (Table 1).

3. Examination of gene and protein expression profile of proteases in human milk.

Previous work has suggested that plasmin is found in milk but that it originates from blood (*20*). This study tested if the enzymes responsible for the observed peptide fragments are produced in the mammary gland or migrate into milk from other origins. Gene expression analysis was carried out for human milk according to the procedure used for bovine milk ((*15*); Table 1). The expression was analyzed for days 4, 15, 30 and 60 of lactation. For consistency purposes, day 60 gene expression was determined, however we also took into account the expression levels for the other days of lactation to build Table 1 which shows the possibility of expression of an enzyme at different stages of lactation. The results show that very few of the proteolytic enzymes described in Table 1 are expressed in the mammary

gland at these timepoints (Table 1). In agreement with previous literature, no gene expression of plasmin was detected in the milks. Moreover, no expression was observed for the intestinal enzyme trypsin.

Cathepsin D was highly expressed (365 RPKM, Table 1). Elastase, however, was not expressed in the mammary epithelial cells (Table 1). As mentioned above, the presence of active elastase in milk may be explained by its presence in PMNs. We would expect the vast majority of PMN cells to be pelleted by the centrifugation and thus, would not participate in any degradation of the milk proteins after centrifugation (see methods). However, PMN can secrete elastase while still within the mammary gland, and the predicted elastase activity may derive from them.

4. Other predicted proteases that were also expressed in human milk

Most of the other predicted enzymes have a low number of total cleavages, making it difficult to fully support their presence in milk (Table 1; total cleavages<=44). Interestingly, some of these enzymes activities predicted in milk seem to be expressed in matching proteases. Based on a relatively specific cleavage site not shared by other proteases in these analyses (cleaves after E; with 76 cleavages in total), we predicted a glutamyl endopeptidase-like activity. Transcripts for glutamyl endopeptidase were, however, not detected in mammary epithelial cells. A glutamyl endopeptidase-like protease (proteasome subunit beta type-3, PSMB3) is, however, highly expressed (75.9 RPKM; Table 1). This result may explain the high number of cleavage sites predicted to be due to a glutamyl endopeptidase-like enzyme (930). Indeed, this glutamyl endopeptidase-like protease may have a more specific cleavage pattern explaining the many residues containing glutamic acid (E) that it did not cleave, as this cannot be accounted for by a structural bias, since the charged glutamic acids are not found buried in the core of the proteins. While the gene expression of a

329 glutamyl endopeptidase (called “a disintegrin and metalloproteinase with
330 thrombospondin motifs 4” (ADAMTS4) was detected, its potential cleavage site,
331 E'[AS], is only cleaved eight times in the dataset. Thus, there is no indication for this
332 enzyme playing a role as the major glutamyl endopeptidase activity in milk.
333 Proline-endopeptidase was found to be expressed in milk (5.52 RPKM; Table
334 1). Proline-endopeptidase is responsible for the cleavage of 45 residues covering a
335 total of 18 milk proteins (Table 1; Figure 1).
336 **5. *In vivo* release of antimicrobial peptides in human milk**
337 Milk proteins may carry encrypted functional peptide sequences that, when
338 released by enzymes from the intact protein, help in the protection and development
339 of the neonate (*12*). But without *in vivo* results, these concepts remain unproven. This
340 study identified four peptides that overlap (2 extra or less amino acids on the *N*-
341 terminus of the peptide) with a known antimicrobial peptide that has been reported
342 previously in the literature (*21*). This peptide is present at the *C*-terminus of β-casein
343 (Figure 2). These four overlapping peptides are naturally released in human milk via
344 the action of several enzymes, including cathepsin D and elastase (Figure 2). The
345 overlapping peptides most likely carry the antibacterial activity, as the amino acid
346 additions compared to the literature-defined sequence is unlikely to abolish the
347 antimicrobial activity of these sequences.
348 **6. Uneven distribution of enzyme activity in human milk**
349 To measure the enzyme activity we considered the unique cleavages of each
350 (i.e. if an enzyme cleaves out the same peptide from the parent protein multiple times,
351 we will only consider this to be one unique cleavage). This measure highlights more
352 the range of cleavages of an enzyme and makes it possible to compare it with other
353 enzyme ranges rather than comparing its ability to cleave many times the same

peptides. Consequently, this measure does not take the abundance of the proteins into account. Using this approach we find that two proteins—β-casein and polymeric immunoglobulin receptor—show the highest susceptibility to the milk enzyme activity, as shown by the large number of fragments found from these two proteins. β-casein is a milk-specific protein expressed during lactation. Interestingly, the other two milk-specific proteins—α_{S1}-casein and κ-casein—show little digestion susceptibly within the mammary gland. The possibility of protein structural disorder being the source of variability in degradation susceptibility among these proteins is not supported by these results, as κ-casein and α_{S1}-casein are also highly disordered proteins. The K and R residues are those that show the most susceptibility for the cleavages observed in human milk proteins. However, no enrichment of these amino acids was found in β-casein versus α_{S1}-casein or κ-casein (14 K and R sites in β-casein; 16 in α_{S1}-casein; and 13 κ-casein). In fact, there are even less K and R in β-casein over the sequence length compared to α_{S1}-casein or κ-casein, arguing that this is not the driver behind β-casein susceptibility, nor is it for polymeric immunoglobulin receptor (6% in β-casein; 9% in α-S1-casein; 8% κ-casein; and 10% in polymeric immunoglobulin receptor). On the other hand, post-translational modifications may also explain these variations in susceptibility of proteins to degradation. Indeed, large modifications on proteins may prevent enzymes from reaching their target site. However, investigating the potential modifications sites using Uniprot shows that all the casein proteins are slightly/equivalently modified (in terms of numbers of glycosylations and phosphorylations).

The other major milk-specific proteins that have not been affected are α-lactalbumin, and lactoferrin. No peptides from these proteins were detected despite the fact that plasmin, trypsin, cathepsin D and elastase were all predicted to cleave

both of these proteins. It is perhaps because of the particular structure and post-translational modifications of α-lactalbumin, and lactoferrin that prevented the enzymes from cleaving them.

Although cathepsin D is highly expressed in milk (Table 1), it does not show a high effectiveness in cleaving milk proteins. This lower cleavage rate may be because cathepsin D is most effective at a more acidic pH (pH=5 for cathepsin D), which is different to the more neutral milk pH.

Independent studies have shown the existence and activity of some enzymes in milk, but the extent to which these enzymes are active relative to each other remained unknown. Likewise, whether or not these enzymes are created in the mammary epithelial cells or not was unclear. The enzymes that are active in human milk and responsible for protein digestion prior to entry into the infant's digestive system were determined. Using a combination of bioinformatics, peptidomics, transcriptomics and literature results, we showed that milk begins to be digested even before infant consumption, and that this digestion is mainly carried out by four proteolytic enzymes: plasmin, a trypsin-like enzyme, elastase, and cathepsin D (Tables 1, 2, 3, 5, 6). Among these four enzymes, only cathepsin D and elastase are expressed in milk, with elastase showing only weak expression at day 15 and 30, and no expression after that (Table 1). Activity and expression of chymotrypsin and pepsin-like activity were also seen (Table 1). This finding is surprising, given that these enzymes have never been reported in human breast milk. Other enzymes that this analysis predicted from activity and expression are glutamyl endopeptidase and proline endopeptidase (Table 1).

Interestingly, plasmin, trypsin, elastase and cathepsin D cleave the casein proteins (α_{s1}-, β-, and κ-casein) to various extents. Our novel MS-based peptide

search platform using an iterative searching strategy to detect peptides in minor abundances did not detect peptides resulting from α-lactalbumin. Plasmin, a trypsin-like enzyme, elastase and cathepsin are predicted to cleave this protein, yet none of the predicted peptides could be detected. The mechanism causing this protein selectivity of digestion remains unknown.

We know very little about proteolytic enzyme activity in human milk, this makes the problem of humanizing formulae milk even harder. The work we have carried out here sheds light on the different enzymes we find in human milk and the relative contribution of each of them –in terms of the release of unique peptides- to cleaving milk proteins. The example illustrated in figure 2 of a known anti-microbial activity being released illustrates how enzyme activity in human milk may be playing a much greater role than previously anticipated, and fully clarifying this role is key in understanding the infant needs.

ABBREVIATIONS

MS, mass spectrometry; FA, formic acid; ACN, acetonitrile; OD, odds ratio; PMN, polymorphonuclear neutrophils; RPKM, reads per kilobase per million mapped reads.

ACKNOWLEDGEMENTS

The authors thank Prof. Alan Kelly and Kevin Rue for helpful discussions.

This work was funded by the Irish Research Council for Science, Engineering and Technology, co-funded by Marie Curie Actions under FP7; Enterprise Ireland grant to Food Health Ireland; **SFI** grant 08/IN.1/B1864; University of California Discovery Program (05GEB01NHB), the National Institute of Health (HD059127 and

HD061923) California Dairy Research Foundation; United States Department of Agriculture, National Institute for Food and Agriculture post-doctoral fellowship; National Science Foundation Graduate Research Fellowship Program; National Institute of Health Training Program in Biomolecular Technology (2-T3-GM08799).

Notes

The authors declare no competing financial interest.

REFERENCES

1. Heegaard, C. W.; Larsen, L. B.; Rasmussen, L. K.; Højberg, K. E.; Petersen, T. E.; Andreasen, P. A., Plasminogen activation system in human milk. *J. Pediatr. Gastroenterol. Nutr.* **1997,** *25*, 159-166.

2. Nielsen, S. S., Plasmin System and Microbial Proteases in Milk: Characteristics, Roles, and Relationship. *J. Agric. Food Chem.* **2002,** *50*, 6628-6634.

3. Borulf, S.; Lindberg, T.; Mansson, M., Immunoreactive anionic trypsin and anionic elastase in human-milk. *Acta Paediatr.* **1987,** *76*, 11-15.

4. Monti, J.; Mermoud, A.; Jolles, P., Trypsin in human milk. *Experientia* **1986,** *42*, 39-41.

5. Kaminogawa, S.; Yamauchi, K., Acid protease of bovine milk. *Agric. Biol. Chem.* **1972,** *36*, 2351-2356.

6. Armaforte, E.; Curran, E.; Huppertz, T.; Ryan, C. A.; Caboni, M. F. O.; Connor, P. M.; Ross, R. P.; Hirtz, C.; Sommerer, N.; Chevalier, F., Proteins and proteolysis in pre-term and term human milk and possible implications for infant formulae. *Int. Dairy J.* **2010,** *20*, 715–723.

7. Ferranti, P.; Traisci, M. V.; Picariello, G.; Nasi, A.; Boschi, V.; Siervo, M.; Falconi, C.; Chianese, L.; Addeo, F., Casein proteolysis in human milk: tracing the

455 pattern of casein breakdown and the formation of potential bioactive peptides. *The*
456 *Journal of dairy research* **2004,** *71*, 74-87.
457 8. Borulf, S.; Lindberg, T.; Mansson, M., Immunoreactive anionic trypsin and
458 anionic elastase in human milk. *Acta paediatrica Scandinavica* **1987,** *76*, 11-5.
459 9. Monti, J. C.; Mermoud, A. F.; Jolles, P., Trypsin in human milk. *Experientia*
460 **1986,** *42*, 39-41.
461 10. Shahani, K.; Kwan, A.; Friend, B. A., Role and significance of enzymes in
462 human milk. *Am. J. Clin. Nutr.* **1980,** *33*, 1861-1868.
463 11. Korhonen, H.; Pihlanto, A., Bioactive peptides: production and functionality.
464 *Int. Dairy J.* **2006,** *16*, 945-960.
465 12. Dallas, D. C.; Underwood, M. A.; Zivkovic, A. M.; German, J. B., Digestion
466 of protein in premature and term infants. *Journal of Nutritional Disorders and*
467 *Therapy* **2012,** *2*, 1-9.
468 13. Dallas, D. C.; Guerrero, A.; Khaldi, N.; Castillo, P. A.; Martin, W. F.;
469 Smilowitz, J. T.; Bevins, C. L.; Barile, D.; German, J. B.; Lebrilla, C. B., Extensive in
470 vivo human milk peptidomics reveals specific proteolysis yielding protective
471 antimicrobial peptides. *J. Proteome Res.* **2013,** *12*, 2295-2304.
472 14. Vijayakumar, V.; Guerrero, A.; Davey, N.; Lebrilla, C. B.; Shields, D. C.;
473 Khaldi, N., EnzymePredictor: a tool for predicting and visualizing enzymatic
474 cleavages of digested proteins. *J. Proteome Res.* **2012,** *11*, 6056-6065.
475 15. Wickramasinghe, S.; Rincon, G.; Islas-Trejo, A.; Medrano, J. F.,
476 Transcriptional profiling of bovine milk using RNA sequencing. *BMC Genomics*
477 **2012,** *13*, 45-58.
478 16. Mortazavi, A.; Williams, B. A.; McCue, K.; Schaeffer, L.; Wold, B., Mapping
479 and quantifying mammalian transcriptomes by RNA-Seq. *Nature methods* **2008,** *5*,
480 621-628.
481 17. Larsen, L. B.; Petersen, T. E., Identification of five molecular forms of
482 cathepsin D in bovine milk. In *Aspartic Proteinases*, Springer: 1995; pp 279-283.

483 18. Travis, J.; Fritz, H., Potential problems in designing elastase inhibitors for
484 therapy. *Am. Rev. Respir. Dis.* **1991,** *143*, 1412-1415.
485 19. Prin-Mathieu, C.; Le Roux, Y.; Faure, G.; Laurent, F.; Béné, M.; Moussaoui,
486 F., Enzymatic activities of bovine peripheral blood leukocytes and milk
487 polymorphonuclear neutrophils during intramammary inflammation caused by
488 lipopolysaccharide. *Clin. Diagn. Lab. Immunol.* **2002,** *9*, 812-817.
489 20. Kelly, A.; O'Flaherty, F.; Fox, P., Indigenous proteolytic enzymes in milk: A
490 brief overview of the present state of knowledge. *Int. Dairy J.* **2006,** *16*, 563-572.
491 21. Hayes, M.; Stanton, C.; Fitzgerald, G. F.; Ross, R. P., Putting microbes to
492 work: Dairy fermentation, cell factories and bioactive peptides. Part II: Bioactive
493 peptide functions. *Biotechnology journal* **2007,** *2*, 435-449.
494 22. Wilkins, M. R.; Gasteiger, E.; Bairoch, A.; Sanchez, J.-C.; Williams, K. L.;
495 Appel, R. D.; Hochstrasser, D. F., Protein identification and analysis tools in the
496 ExPASy server. In *2-D Proteome Analysis Protocols*, Springer: 1999; pp 531-552.
497 23. Christensen, B.; Schack, L.; Kläning, E.; Sørensen, E. S., Osteopontin is
498 cleaved at multiple sites close to its integrin-binding motifs in milk and is a novel
499 substrate for plasmin and cathepsin D. *J. Biol. Chem.* **2010,** *285*, 7929-7937.

500
501

Figure Captions

Figure 1. Representation of the total number of cleaved milk proteins per predicted enzyme. Each enzyme name on the X-axis is plotted against the total number of proteins it cleaves on the Y-axis.

Figure 2. Illustration of antimicrobial-like peptides released *in vivo* from β-casein in human milk. The peptide QELLLNPTHQIYPVTQPLAPVHNPISV shown at the top of the figure in grey has been discovered to be an antimicrobial peptide (*21*). This peptide is found at the *C*-terminus of β-casein (position 199 to 225). Four overlapping peptides found to be released *in vivo* in human milk are represented under the β-casein sequence. The *in vivo* cleavage positions are indicated with grey lines. The enzymes predicted to be responsible for these cleavages are represented beside the grey lines.

TABLES

Table 1. Details of the enzymes that take part in the digestion of the human milk proteins ranked based on their total number of cleavages. The corresponding gene expression levels in RPKM are provided for each enzyme.

Enzymes	*N*-terminus cleavage count	*C*-terminus cleavage count	Total cleavage	Unique cleavage	Number of expected cleavages within the peptide	Number of proteins cleaved	Odds ratio	Std error	Gene expression given in RPKM
Plasmin	120	200	320	0	371	15	4.11	1.08	0
Trypsin 1*	120	199	319	0	352	14	4.33	1.09	0 (5.81 PRKM for trypsin domain containing protease TYSND1)
Cathepsin D	68	62	130	0	846	24	0.61	1.1	365.07
Chymotrypsin low 1	**68**	**34**	**102**	**0**	**449**	**11**	**0.93**	**1.12**	**0**
Elastase	**51**	**39**	**90**	**0**	**563**	**13**	**0.64**	**1.12**	**0**
Pepsin 1 (pH1.3)	31	47	78	0	479	10	0.66	1.13	0.1

Glutamyl endopeptidase*	55	21	76	0	930	5	0.31	1.13	0 (75.97 PRKM for proteasome subunit beta type-6, responsible for the peptidyl glutamyl-like activity, PSMB6)
Pepsin 1 (pH>2)	30	38	68	0	315	8	0.89	1.15	0.1
Proline endopeptidase	14	31	45	0	598	18	0.29	1.17	5.52
Chymotrypsin low 4	**3**	**12**	**15**	**0**	**102**	**4**	**0.60**	**1.32**	**0**
Chymotrypsin low 3	**6**	**6**	**12**	**0**	**79**	**6**	**0.62**	**1.36**	**0**
Chymotrypsin low 2	**2**	**0**	**2**	**0**	**0**	**1**	**20.65**	**4.71**	**0**

Enzymes in bold represent ones that are not expressed at day 60 of lactation but low expression of these genes has been detected at earlier stages of lactation (15, and 30 days).

* This indicates enzymes that do not show expression in milk but expression data shows the presence of other enzymes that have a similar activity and are expressed.

Table 2. Human milk proteins found to have been digested by plasmin. The proteins are ordered based on the number of cleavages plasmin has performed on each. Here we only considered the number of peptides that are unique, if a peptide was not unique we consider only one copy of the peptide. Proteins exclusively found in milk are represented in bold.

Protein name	Uniprot ID and access name	Total number of cleavages performed
β-casein	P05814 (CASB_HUMAN)	89
Polymeric immunoglobulin receptor	P01833 (PIGR_HUMAN)	32
Osteopontin	P10451 (OSTP_HUMAN)	18
Butyrophilin subfamily 1 member A1	Q13410 (BT1A1_HUMAN)	15
α-S1-casein	P47710 (CASA1_HUMAN)	11
K-casein	P07498 (CASK_HUMAN)	7
Mucin-1	P15941 (MUC1_HUMAN)	3
Parathyroid hormone-related protein	P12272 (PTHR_HUMAN)	2
Bile salt-activated lipase	P19835 (CEL_HUMAN)	2
Androgen receptor	Q9UN21 (Q9UN21_HUMAN)	1
Protein diaphanous homolog 1	O60610 (DIAP1_HUMAN)	1
Complement C4-A	P0C0L4 (CO4A_HUMAN)	1

La-related protein 1	Q6PKG0 (LARP1_HUMAN)	1
NMDA receptor-regulated protein 2	Q659A1 (NARG2_HUMAN)	1
Dedicator of cytokinesis protein 1	Q14185 DOCK1_HUMAN	1

Table 3. Human milk proteins found to have been digested by trypsin. The proteins are ordered based on the number of cleavages trypsin has performed on each. Here we only considered the number of peptides that are unique, if a peptide was not unique we consider only one copy of the peptide. Proteins exclusively found in milk are represented in bold.

Protein name	Uniprot ID and access name	Total number of cleavages performed
β-casein	P05814 (CASB_HUMAN)	89
Polymeric immunoglobulin receptor	P01833 (PIGR_HUMAN)	32
Osteopontin	P10451 (OSTP_HUMAN)	18
Butyrophilin subfamily 1 member A1	Q13410 (BT1A1_HUMAN)	15
α_{S1}-casein	P47710 (CASA1_HUMAN)	11
κ-casein	P07498 (CASK_HUMAN)	7
Mucin-1	P15941 (MUC1_HUMAN)	3
Parathyroid hormone-related protein	P12272 (PTHR_HUMAN)	2
Bile salt-activated lipase	P19835 (CEL_HUMAN)	2
La-related protein 1	Q6PKG0 (LARP1_HUMAN)	1
Protein diaphanous homolog 1	O60610 (DIAP1_HUMAN)	1
NMDA receptor-regulated	Q659A1 (NARG2_HUMAN)	1

protein 2		
Complement C4-A	P0C0L4 (CO4A_HUMAN)	1
Dedicator of cytokinesis protein 1	Q14185 (DOCK1_HUMAN)	1

Table 4. Enzyme specificity for plasmin and trypsin.

Enzyme	Cleavage Pattern						Reference
	P4	P3	P2	P1	P1’	P2’	
Trypsin	-	-	W	K	P	-	Wilkins et al., 1999)
	-	-	M	R	P	-	(*22*)
Plasmin	-	-	-	K or R	-	-	(*23*)

Table 5. Human milk proteins found to have been digested by cathepsin D. The proteins are ordered based on the number of cleavages trypsin has performed on each. Here we only considered the number of peptides that are unique, if a peptide was not unique we consider only one copy of the peptide. Proteins exclusively found in milk are represented in bold.

Protein name	Uniprot ID and access name	Total number of cleavages performed
β-casein	P05814 (CASB_HUMAN)	78
Polymeric immunoglobulin receptor	P01833 (PIGR_HUMAN)	21
Perilipin-2	Q99541 (PLIN2_HUMAN)	3
Osteopontin	P10451 (OSTP_HUMAN)	2
α_{S1}-casein	P47710 (CASA1_HUMAN)	2
Deubiquitinating protein VCIP135	Q96JH7 (VCIP1_HUMAN)	2
Butyrophilin subfamily 1 member A1	Q13410 (BT1A1_HUMAN)	2
Receptor-type tyrosine-protein phosphatase	Q15262 (PTPRK_HUMAN)	2
Macrophage mannose receptor 1	P22897 (MRC1_HUMAN)	2
Protein CASC3	O15234 (CASC3_HUMAN)	1
Flavin containing monooxygenase 5, isoform CRA_c	Q9HA79 (Q9HA79_HUMAN)	1
Abl interactor 1	Q8IZP0 (ABI1_HUMAN)	1
Misshapen-like kinase 1	Q8N4C8 (MINK1_HUMAN)	1

La-related protein 1	Q6PKG0 (LARP1_HUMAN)	1
Receptor-type tyrosine-protein phosphatase α	P18433 (PTPRA_HUMAN)	1
Ubiquitin carboxyl-terminal hydrolase 51	Q70EK9 (UBP51_HUMAN)	1
Protein diaphanous homolog 1	O60610 (DIAP1_HUMAN)	1
Neural Wiskott-Aldrich syndrome protein	O00401 (WASL_HUMAN)	1
κ-casein	P07498 (CASK_HUMAN)	1
Insulin receptor substrate 1	P35568 (IRS1_HUMAN)	1
Transcription factor 7-like 2	Q9NQB0 (TF7L2_HUMAN)	1
Gamma-glutamyltransferase 6	Q6P531 (GGT6_HUMAN)	1
PH domain leucine-rich repeat-containing protein phosphatase 1	O60346 (PHLP1_HUMAN)	1
Dedicator of cytokinesis protein 1	Q14185 (DOCK1_HUMAN)	1

Table 6. Human milk proteins found to have been digested by elastase. The proteins are ordered based on the number of cleavages trypsin has performed on each. Here we only considered the number of peptides that are unique, if a peptide was not unique we consider only one copy of the peptide. Proteins exclusively found in milk are represented in bold.

Protein name	Uniprot ID and access name	Total number of cleavages performed
β-casein	P05814 (CASB_HUMAN)	41
Polymeric immunoglobulin receptor	P01833 (PIGR_HUMAN)	25
Butyrophilin subfamily 1 member A1	Q13410 (BT1A1_HUMAN)	8
α_{S1}-casein	P47710 (CASA1_HUMAN)	3
Perilipin-2	Q99541 (PLIN2_HUMAN)	3
Protein CASC3	O15234 (CASC3_HUMAN)	1
Androgen receptor	Q9UN21 (Q9UN21_HUMAN)	1
Parathyroid hormone-related protein	P12272 (PTHR_HUMAN)	1
Osteopontin	P10451 (OSTP_HUMAN)	1
Heat shock protein β-1	P04792 (HSPB1_HUMAN)	1
Receptor-type tyrosine-protein phosphatase K	Q15262 (PTPRK_HUMAN)	1
Receptor-type tyrosine-protein phosphatase α	P18433 (PTPRA_HUMAN)	1
Gamma-glutamyltransferase 6	Q6P531 (GGT6_HUMAN)	1

FIGURES

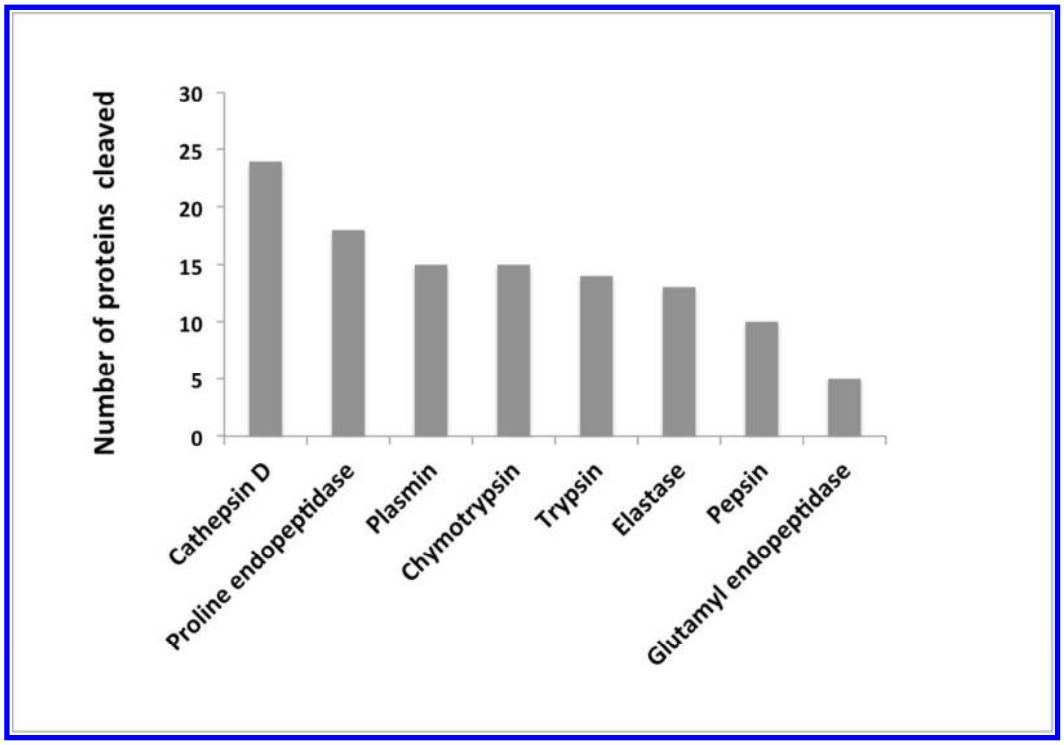

Figure 1.

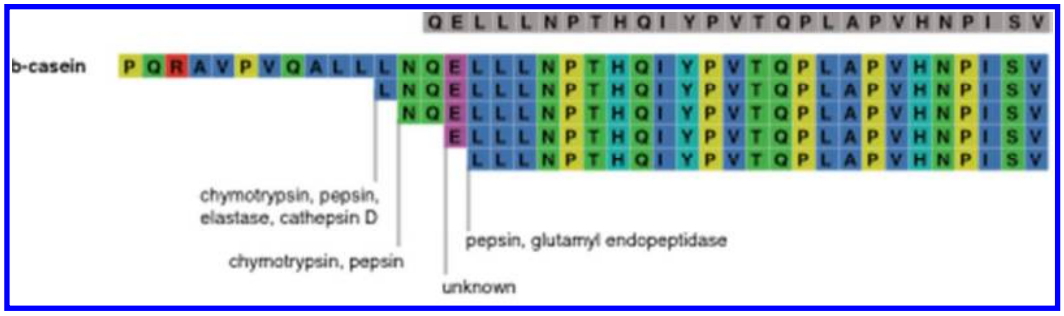
b-casein
chymotrypsin, pepsin, elastase, cathepsin D
chymotrypsin, pepsin
pepsin, glutamyl endopeptidase
unknown

Figure 2.

TOC Graphic

Graphic for Table of Contents