

## **Endogenous antimicrobial peptides in human colostrum: a peptidomics study by liquid-chromatography high-resolution mass-spectrometry (LC-HRMS) and bioinformatics**

Isabele B. Campanhon<sup>a,b</sup>, Nicole Cavalcante da Silva<sup>a,b</sup>, Rosane de Oliveira Nunes<sup>a</sup>,  
Flávia Fioruci Bezerra<sup>c</sup>, Alexandre Guedes Torres<sup>a,b,‡,\*</sup>, Márcia R. Soares<sup>a,‡</sup>

<sup>‡</sup> These authors share senior authorship

Authors' affiliations:

<sup>a</sup> Department of Biochemistry, Institute of Chemistry, Universidade Federal do Rio de Janeiro, Brazil

<sup>b</sup> Lipid Biochemistry and Lipidomics Laboratory and Laboratory of Food Science and Nutritional Biochemistry, Institute of Chemistry, Universidade Federal do Rio de Janeiro, Brazil

<sup>c</sup> Department of Basic and Experimental Nutrition, Nutrition Institute, State University of Rio de Janeiro, Brazil

\*Corresponding Author: [torres@iq.ufrj.br](mailto:torres@iq.ufrj.br)

Av. Horácio Macedo, 1281, lab. 245 - Pólo de Química

Cidade Universitária, Ilha do Fundão

CEP: 21941-598

Rio de Janeiro – RJ

**Competing Interests:** The authors have no competing interests to declare.

List of abbreviations: AMP, antimicrobial peptide; COVID-19, coronavirus disease 2019; HIV, human immunodeficiency virus; LC-HRMS, liquid-chromatography high-resolution mass-spectrometry; WHO, World Health Organization.

## **Abstract**

Human colostrum is recognized as a source of bioactive compounds that impact on neonatal development, such as antimicrobial peptides. Antimicrobial peptides (AMPs) present some common characteristics, as they are relatively small (<10 kDa), most of them are positively charged and possess  $\alpha$ -helical regions. These molecules can protect the newborn against infectious agents such as bacteria, fungi, and viruses. Using an MS-based proteomic approach combined with the application of bioinformatic tools, 29 peptide sequences from human colostrum were annotated in human colostrum. These peptides are potentially antifungal and were enriched in glutamic acid (E), phenylalanine (F), lysine (K) and tryptophan (W) residues. Among the precursor proteins, two have a known 3D structure and their fragments are located in the protein core. In addition, *in silico* analyses revealed that AMPs containing regions of  $\alpha$ -helix and random structures, and high aliphatic index (mean of 85.29) tend to be thermostable. In this work with human colostrum, previously unreported peptides with potential antimicrobial activity were annotated and structurally characterized via bioinformatic tools.

**Keywords:** endogenous peptides; antimicrobial peptides; bioinformatics

## 1. Introduction

The main causes of neonatal death include infections, prematurity and birth asphyxia, and the protective role of breast milk against infections has been widely reported [1-3]. Protective factors in breast milk include immunoglobulins, and bioactive compounds, such as antihypertensive, antioxidant, antithrombotic and antimicrobial peptides [9]. These factors can also stimulate the maturation of the neonatal immune system [4,5]. Antimicrobial peptides (AMPs) were active against several pathogens [10] and thus they might reduce neonatal mortality by infections. Endogenous peptides can be released via the action of milk proteases such as plasmin, cathepsin D and elastase, and proteolytic activity can be modulated by activators and inhibitors secreted by mammary epithelial cells [6]. Some of these endogenous proteases may be active in the mammary alveoli before milk ejection [7], and the extent of proteolytic activity between these proteases might differ in intact milk and during infant digestion [8].

Antimicrobial peptides in breast milk might protect the newborn against emerging viral diseases, such as COVID-19, and to bacteria resistant to antimicrobials [11]. In addition, the prospection for endogenous bioactive peptides may contribute to future developments of infant formulas that would help to prevent neonatal infections [12]. As most studies on bioactive peptides in breast milk investigated those formed during *in vitro* digestion of mature milk [13,14], endogenous peptides in colostrum remain largely unresearched. Previous studies showed that the profile of endogenous peptides in milk varied between women delivering at term and those delivering prematurely, and peptides in these groups were expressed differently [15,16]. In this work, colostrum ultrafiltration followed by nano-liquid-chromatography coupled to high-resolution mass spectrometry (nLC-Orbitrap-HRMS), combined with bioinformatic tools were used to annotate and characterize the endogenous peptides of human colostrum with potential antifungal activity. Optimization of antimicrobial peptide structures might support future clinical applications, in addition to being a natural source for the discovery of new antimicrobial peptides.

## 2. Materials and methods

### 2.1. Study volunteers and breast milk collection

Mothers who delivered at term were recruited, and colostrum samples were obtained from twenty-four healthy adult volunteers. Samples were collected by manual expression into polypropylene containers at the Herculano Pinheiro Maternity Hospital in Rio de Janeiro, which is a public health service unit. All donors were between the ages of 21 and 40 years with no diagnosed chronic diseases, and were not taking any prescribed medications or vitamin supplements. Participants were recruited between 37 to 41 weeks of pregnancy, and colostrum samples were collected between 1 and 5 days postpartum by standard procedures to avoid microbial contamination. The research protocol was approved by the Ethical Committee of Maternity School of the Federal University of Rio de Janeiro (protocol number CAAE: 56617516.5.0000.5259), and the mothers signed informed consent before participating in the study. Milk samples were transported on ice to the laboratory and stored at  $-80\text{ }^{\circ}\text{C}$  until analysis. Maternal and neonatal characteristics are presented in **Supplementary Table 1**.

### 2.2. Sample preparation

Colostrum samples were thawed in a  $37\text{ }^{\circ}\text{C}$  water bath and protein concentration was determined according to the Lowry [17] protein assay. Subsequently, a 5.0 mL pooled sample was prepared, containing all 24 samples balanced for their protein contents so that the pooled sample had the same amount of protein from each donor milk.

Milk fat was removed by centrifuging ( $1,500\times g$ , 20 min) and collecting the defatted aqueous fraction consisting of skim milk [18]. The skim milk was collected and ultra-filtered through Microcon<sup>®</sup> concentrators (YM-10K, Millipore, Billerica, MA, USA), at  $14,000\times g$  for 15 minutes at  $10\text{ }^{\circ}\text{C}$  to separate the peptide ( $< 10\text{ kDa}$ , filtrate) and protein ( $\geq 10\text{ kDa}$ , retained) fractions.

### 2.3. Mass spectrometry analysis and bioinformatics

The samples were analyzed in triplicate in a nano-LC Ultimate 3000 (Thermo Fisher Scientific, IL, USA) coupled to a quadrupole-Orbitrap mass analyzer platform (Q-Exactive Plus; Thermo Fisher Scientific, Bremen, GER). An aliquot containing  $1\text{ }\mu\text{g}$  of peptides was loaded on a guard-column (2 cm length,  $200\text{ }\mu\text{m}$  i.d.) packed with

ReproSil-Pur C18-AQ (5  $\mu\text{m}$  particle diameter) and separated in an analytical column Picochip with ReproSil-Pur C18 (3  $\mu\text{m}$  particle diameter). Peptides were eluted using a gradient of 95% solvent A (95%  $\text{H}_2\text{O}$ , 5% acetonitrile, 0.1% formic acid) to 40% solvent B (95% acetonitrile, 5%  $\text{H}_2\text{O}$ , 0.1% formic acid) for 60 min; 40% to 85% of solvent B for 10 min; and 95% of solvent B for 15 min, at a 300 nL/min flow rate.

Full scan and MS/MS were acquired on positive mode using a data-dependent acquisition (DDA; dd-MS<sup>2</sup>) protocol. The ion source and S-lens were optimized to spray voltage at 3.4 kV, zero flow of sheath and auxiliary gas at 250 °C and 80 S-Lens RF level. MS full scan was acquired at 70,000 mass resolution scanning from 400 to 2000 m/z in the Orbitrap analyzer, 10<sup>6</sup> AGC, and 50 ms maximum ion injection time. The 10 most intense ions containing 2 to 4 charges were selected for high-energy collision dissociation (HCD) fragmentation. Fragment acquisition in the Orbitrap using 30 a.u. normalized collision energy, and dynamic exclusion was enabled for 20 s. The MS/MS scans were obtained using 17,500 resolution,  $5 \times 10^4$  AGC, 50 ms maximum isolation time, 1 sec microscans, 200-2000 m/z range, and 2.0 m/z isolation window.

The DDA raw data were processed and searched by the Proteome Discovery 2.1 software server search engine (Thermo Fisher Scientific, Bremen, GER) with the SEQUEST algorithm by tolerating up to  $\pm 0.1$  Da for precursor ions and 0.01 Da for fragment ions. Protein identification was performed by searching the mass spectrometric data against the UniProt-SwissProt protein database (available in April 2018) containing reversed sequences with a false discovery rate (FDR) < 1 %.

#### *2.4. Search for functional endogenous peptides*

Peptide sequences obtained by mass spectrometry were compared to sequences from four functional peptide databases: Milk bioactive peptide database [19], AMPA (Antimicrobial Sequence Scanning System) [20], ClassAMP (A prediction tool for classification of AMPs) [21] and APD (The antimicrobial peptide database) [22]. The milk bioactive peptide database was constructed from peptides derived from milk proteins described in previous research articles with a biological function [19]. The AMPA tool is based on the characteristics of each amino acid and is used to classify different protein sequences with antimicrobial or non-antimicrobial regions. ClassAMP was used to predict the propensity of a peptide sequence to present antibacterial, antifungal or antiviral activity. APD was used to assess the similarity between the

sequences annotated and those deposited in public libraries, and only the peptides with at least 40% similarity were considered.

### 2.5. *In silico analysis of the three-dimensional structure of proteins and peptides*

The three-dimensional structure of the precursor proteins was obtained from The Protein Data Bank (PDB), and PyMOL software [26] was used to import PDB files. For the peptides sequences, the three-dimensional structure was predicted by the software PEPFOLD3, an online resource for *de novo* peptide structure prediction [27].

### 2.6. *Physicochemical properties of peptides*

Molecular mass, isoelectric point, the number and composition of amino acid residues, values of the instability index, aliphatic index and GRAVY index (grand average of hydrophobicity) of antimicrobial peptides were determined *in silico* by ExPASy-ProtParam (<http://web.expasy.org/protparam>). In addition, the amino acid frequencies were calculated by Genscript software (<https://www.genscript.com/tools/codon-frequency-table>).

### 2.7. *Evaluation of the antimicrobial activity*

Peptide sequences were classified as best AMP candidates if they were matched by at least three bioinformatic tools. Two of the most promising peptides were synthesized commercially and used to assess antimicrobial potential. *Escherichia coli* strain DH5 $\alpha$  was purchased from New England Biolabs (NEB). The antimicrobial activity of the two peptides: AGLAPYKLRPV derived from lactoferrin and HLPLPLLQPL derived from beta-casein was evaluated on batch cultures containing 0.001, 0.01 and 0.1 mg/mL—against *E. coli* DH5 $\alpha$  in a cell suspension of 10<sup>8</sup> CFU/mL grown in LB media (Luria-Bertani) (BD™, Le Pont de Claix, France). Cells were then serially diluted 10-fold each step in 8.5 g/L NaCl, 2.0 g/L Tween 80, plated on solid LB and after incubation at 37 °C for 18 h the colony-forming units were counted. The procedure was performed in triplicate, and inhibition was calculated as described elsewhere [28].

## 2.8. Statistical analysis

Data were expressed as mean  $\pm$  standard deviation, and were processed using Prism for Windows software, version 5.04 (GraphPad Software, San Diego, CA). Paired *t*-tests were used to test differences between peptide sequences. A *p*-value  $< 0.05$  was considered as statistically significant.

## 3. Results

### 3.1. Maternal and neonatal characteristics

**Supplementary Table 1** shows the characteristics of the colostrum donors and their newborns during sample collection.

### 3.2. Endogenous Peptides analysis

Endogenous peptides were determined in a pooled sample of human colostrum from 24 adult mothers. Colostrum was defatted by centrifugation, the skimmed colostrum was ultrafiltered (through 10 kDa cut-off membranes), and the ultrafiltered fraction was analyzed ( $< 10$  kDa peptides). A total of 29 peptide sequences were identified by Mass Spectrometry (MS/MS) deriving from 15 precursor proteins (**Table 1**). The major milk proteins such as caseins, lactoferrin and immunoglobulins were identified among precursors of the peptide fragments. Most ( $> 40\%$ ) of the peptide sequences were derived from  $\beta$ -casein.

### 3.3. Physicochemical properties of peptides

Physicochemical properties of individual peptides are presented in **Table 2**, showing that some of these properties varied broadly among the annotated peptides. Most of the peptides showed a negative hydropathy index value indicating that these sequences have a hydrophilic nature (**Table 2**). In the two groups of peptides showing either hydrophilic or hydrophobic nature, the isoelectric point was on average 5.5 and 5.7, respectively, indicating a mildly acidic nature. The annotated peptides contained from 9 to 24 amino acid residues, and the frequency of occurrence of each amino acid is shown in **Figure 1**.

Although the amino acids' frequency did not vary significantly (paired *t*-test), the amino acids glutamic acid (E), phenylalanine (F), lysine (K) and tryptophan (W) were more prominent.

The aliphatic index of the sequences ranged from 6.67 to 195, but on average it was high (85.29), which is positively correlated with thermostability. The most stable peptides were IRLDIQGTGQLLF from polymeric immunoglobulin receptor, DISNPTAHENYE from alpha-S1-casein, ALGVLVWAA from PRA1 family protein 2, YPDATDEDITSH from osteopontin and DGQVINETSQHHDDLE from vimentin.

### *3.4. Analysis of peptides with potential antimicrobial activity*

In this study, we searched four peptide databases for the annotated endogenous peptides in colostrum. The sequences identified herein have not been previously deposited in the databases of the bioinformatics service Milk Bioactive Peptide Database [19] and AMPA software [20]. In addition, the peptide sequences identified were tested in the search software ClassAMP for antimicrobial functions, and all the peptides presented potential antimicrobial activity (**Supplementary Table 2**). Among these peptides, 45 % were predicted with potential antifungal activity (**Figure 2**).

Additionally, the APD database was used to assess the similarity between the annotated peptide sequences and those deposited in the bank. APD is based on the physicochemical properties of peptides [22]. Some peptides showed a similarity near 40%, these sequences are presented in **Supplementary Table 3**.

### *3.5. In silico 3D structure prediction of endogenous peptides*

The prediction of peptides three-dimensional structure was performed using the PEPFOLD3 software [27], which allows structure prediction of peptides that have between 9 and 36 amino acids in aqueous solution. Structural models of the 29 peptides obtained by this software are shown in **Figure 3**.

### *3.6. 3D structural analysis of precursor proteins*

The structures of the colostrum proteins that were precursors of endogenous peptides were visualized using PyMOL software [26]. Three precursor proteins of endogenous peptides with known structures were shown: the polymeric immunoglobulin receptor



protein, immunoglobulin J and lactoferrin (**Figure 4**). As evident in **Figure 4**, the peptides derived from these proteins emerged from the proteins' cores. The peptide IRLDIQGTGQLLF, derived from the polymeric immunoglobulin receptor protein, has potential antifungal activity predicted by ClassAMP (in blue in **Figure 4**). The peptide AGLAPYKLRPV derived from lactoferrin, has two segments with secondary structure,  $\beta$ -sheet and  $\alpha$ -helix, and it is potentially antiviral (in orange in **Figure 4**). The peptide DPNEDIVERNI from immunoglobulin J has one segment with  $\beta$ -sheet structure and is also potentially antiviral (in pink in **Figure 4**).

### 3.7. Antimicrobial activity

Endogenous peptides in human milk were able to abolish the growth of *E. coli* strain DH5 $\alpha$ . Growth inhibition was achieved after exposure to peptides at 0.01 mg/mL, by 11 % with the peptide HLPLLLQPL (from  $\beta$ -casein) and by 8 % with the peptide AGLAPYKLRPV (from lactoferrin) (**Supplementary Table 4**).

## 4. Discussion

Endogenous peptides result from the action of milk proteases, including plasmin, cathepsin D, and elastase. Human mammary epithelial cells secrete both activators and inhibitors of proteases that could influence the proteolytic activity in milk [6]. Previous studies on milk peptidomics have highlighted the importance of endogenous proteases in the digestion of milk proteins [6, 13,14]. In this work, sample ultrafiltration was combined with high resolution mass spectrometry and bioinformatics tools for the prospection and characterization of endogenous peptides in human colostrum with antimicrobial potential.

Twenty-nine peptide sequences were identified by high resolution mass spectrometry, which were derived from 15 precursor proteins. Several peptides identified in this work were casein-derived fragments via the activity of milk endogenous proteases. Previous studies have shown that human milk caseins are active against several microorganisms, such as *Escherichia coli*, *Helicobacter pylori*, *Listeria* spp., *Salmonella* spp., *Staphylococcus* spp., yeasts, and filamentous fungi [10]. Furthermore, we identified peptide sequences arising from the proteolysis of major proteins in milk whey, such as immunoglobulins and lactoferrin. Currently, substantial efforts are being made to find components of human milk that act in defense against

emerging contagious diseases, and these efforts might prove useful in the control of such diseases, such as COVID-19 [29].

Bioinformatics analysis has been effectively used for the identification of antimicrobial peptides. In this work, four bioinformatics tools were used for the characterization of antimicrobial peptides. The Milk Bioactive Peptide Database (<http://mbpdb.nws.oregonstate.edu>), which includes several peptides derived from a variety of milk products, was searched. For the retrieved bioactive peptides, the AMPA tool was used to calculate the regions in the protein structure that presented potential antimicrobial activity. The potential bioactivity revealed for the endogenous colostrum peptides in this study were not previously described, to the best of our knowledge.

The APD database was searched to determine the potential antimicrobial activity [22] of the annotated peptides in human colostrum. One endogenous peptide in colostrum showed high sequence similarity (>50%) to beta-casein, which is one of the most abundant proteins in human milk. Other fragments showed 40 % similarity or lower to human proteins. However, the analysis showed that all peptides were similar to previously identified antimicrobial peptides.

When the peptide sequences were searched in the ClassAMP database, all 29 annotated sequences were predicted as antimicrobial, mostly antifungal (**Figure 2**). Antifungal peptides act specifically on fungal cells, but some may act against both fungi and bacteria [30,31]. Most studies on antimicrobial peptides emphasize antibacterial activity, rendering antifungal peptides relatively unresearched. In neonates, *Candida* spp. infection occurs in approximately 4 to 18% of ill babies and was the third most common cause of late-onset sepsis [32].

Antimicrobial peptides may present varied amino acid compositions and secondary structures, such as  $\alpha$ -helices,  $\beta$ -sheets, extended helix, and loop structures [33]. For the endogenous peptides in human colostrum, the structure prediction pointed to peptides with secondary structures containing  $\alpha$ -helix and/or random structures, seen in varied arrangements. Most frequently, the AMPs previously described have  $\alpha$ -helix structures [33,34]; therefore, the 12 peptides containing predicted  $\alpha$ -helix structures would be promising candidates for AMPs (**Figure 3**). However, peptides containing random structures have also been previously reported to present antimicrobial activity [34]. Therefore, the peptides NPSESSEPIPL, APQNAVSSEETNDFK, LLNPTHQIYPVTQPLAPVHNPI, and LLNPTHQIYPVTQPLAPVHNPIISV could also exhibit antimicrobial properties. The three precursor proteins with known 3D structures

(polymeric immunoglobulin receptor protein, immunoglobulin J and lactoferrin) revealed that the fragments coming from these proteins originated from the protein core (**Figure 4**). Therefore, we confirmed the importance of protein cleavage to expose these bioactive peptides. These peptides have secondary structures that are similar to those predicted by the Pepfold software. These structural profiles may be associated with antimicrobial activity [34].

The bioactivity of an antimicrobial peptide is strongly related to physical, chemical, and structural properties such as amphipathicity, cationicity, hydrophobicity and  $\alpha$ -helicity [34]. On average, peptide sequences presenting a hydrophobic nature had over 65 % of hydrophobic amino acid residues in their structure (**Figure 1**). This is an important feature of AMPs as it directly influences the peptide's ability to insert into the membranes of microorganisms [35]. Peptides with positive GRAVY values are hydrophobic, hence they favorably interact with the plasma membrane's phospholipid bilayer, which is the first requirement for a peptide to present antimicrobial activity [36].

In all the antimicrobial peptides identified, we observed a high frequency of occurrence of glutamic acid (E), phenylalanine (F), lysine (K) and tryptophan (W) (**Fig. 1**). Additionally, the low values observed for the instability index ( $< 40$ ) for five peptide sequences indicate that these peptides are fairly stable [37], and this is a further essential feature of AMPs. These sequences were derived from the polymeric immunoglobulin receptor protein, alpha S1-casein, osteopontin, vimentin and the PRA1 family protein 2. Among these sequences, we highlight the peptide IRLDIQGTGQLLF derived from the polymeric immunoglobulin receptor protein, which showed the lowest instability index (-3.83). Furthermore, the aliphatic index (AI) refers to the stability of proteins at extreme temperatures. The higher the AI, the more thermostable the protein, but peptides with an aliphatic index higher than 40 are regarded as unstable even when presenting high AI [38]. The results suggest that the endogenous antimicrobial peptides in human colostrum are thermostable, as most sequences showed a high aliphatic index (mean of 85.29). Therefore, these peptides could likely retain bioactivity at human body temperature.

Previous studies have shown that human milk can inhibit the action of several bacteria such as *S. aureus*, *E. coli*, and *P. aeruginosa*, and fungi such as those from the genus *Candida* spp. [39]. The antimicrobial activity of milk and colostrum has been attributed mainly to proteins such as immunoglobulins, lactoferrins, caseins and lysozymes. In the

present study, the low inhibition of *E. coli* may be related to this strain's resistance, as high concentrations of peptides were needed for complete inhibition. The mechanisms of action of antimicrobial peptides vary according to the target microorganism, depending on the physical properties of the membranes and the concentration and three-dimensional structure of the peptides [33]. Future *in vitro* experiments might help to propose mechanisms of action of antimicrobial peptides according to their characteristics and chemical structures.

## 6. Conclusions

Endogenous peptide sequences in human colostrum have common structural features observed in other antimicrobial peptides. From these peptides' profile, assessed by combining nLC-Orbitrap-MS/MS and bioinformatics tools, promising sequence candidates were selected for synthesis, aiming for future confirmation of their antimicrobial activity in *in vitro* assays.

**CRedit Statement, author contributions:** **Isabele Campanhon:** Validation, Formal analysis, Investigation, Writing – Original draft, Visualization; **Nicole Cavalcante:** Formal analysis, Investigation; **Rosane Nunes:** Formal analysis, Investigation; **Flávia Bezerra:** Resources, Writing – Review & editing; **Alexandre Guedes Torres:** Conceptualization, Resources, Writing – Review & editing, Supervision, Funding acquisition; **Márcia Soares:** Conceptualization, Validation, Resources, Data Curation, Writing – Review & editing, Supervision, Funding acquisition. All authors read and approved the final manuscript.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data will be available upon request.

**Conflicts of Interest:** The authors have no conflicts of interest to declare.

## Acknowledgments:

The authors acknowledge the following for funding this work: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, CAPES-Brazil (Finance Code 001); Conselho Nacional de Desenvolvimento Científico e Tecnológico, CNPq-Brazil (Grant # 315579/2021-8), Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro, FAPERJ-Brazil (Grants # E-26/201.161/2021; E-26/010.101016/2018; E-26/010.001436/2019; and E-26/211.280/2021). I.B.C. was a recipient of a CAPES/CNPq PhD scholarship, and A.G.T. was a recipient of CNPq and FAPERJ research fellowships.

## References

1. Van Sadelhoff, J. H., Wiertsema, S. P., Garssen, J., Hogenkamp, A. Free amino acids in human milk: A potential role for glutamine and glutamate in the protection against neonatal allergies and infections. *Front Immunol* 2020; 11: 1007.
2. Gray, K. J., Bordt, E. A., Atyeo, C., Deriso, E., Akinwunmi, B., Young, N et al. Coronavirus disease 2019 vaccine response in pregnant and lactating women: a cohort study. *Am J Obstet Gynecol* 2021; 225(3): 303-e1.
3. Vassilopoulou, E., Feketea, G., Koumbi, L., Mesari, C., Berghea, E. C., Konstantinou, G. N. Breastfeeding and COVID-19: from nutrition to immunity. *Front Immunol* 2021; 12: 661806.
4. Collins, A., Weitkamp, J. H., Wynn, J. L. Why are preterm newborns at increased risk of infection? *Arch Dis Child Fetal Neonatal* 2018; 103(4): F391-F3.
5. Turfkruyer, M., Verhasselt, V. Breast milk and its impact on maturation of the neonatal immune system. *Curr Opin Infect Dis* 2015; 28: 199–206.
6. Dallas, D. C., Smink, C. J.; Robinson, R. C.; Tian, T.; Guerrero, A.; Parker, E. A et al. Endogenous Human Milk Peptide Release Is Greater after Preterm Birth than Term Birth. *J Nutr* 2014; 145(3): 425-433.

7. Khaldi, N., Vijayakumar, V., Dallas, D. C., Guerrero, A., Wickramasinghe, S., Smilowitz, J. T et al. Predicting the important enzymes in human breast milk digestion. *J Agric Food Chem* 2014; 62(29): 7225-7232.
8. Holton, T. A., Vijayakumar, V.; Dallas, D. C.; Guerrero, A.; Borghese, R. A.; Lebrilla, C. B et al. Following the digestion of milk proteins from mother to baby. *J Proteome Res* 2014; 13(12); 5777-5783.
9. Lewis, E. D., Richard, C., Larsen, B. M., Field, C. J. The Importance of Human Milk for Immunity in Preterm Infants. *Clin Perinatol* 2017; 44(1): 23-47.
10. Haque, E., Chand, R. Antihypertensive and Antimicrobial Bioactive Peptides from Milk Proteins. *Eur Food Res Technol* 2008; 227 (1): 7–15.
11. Fox, A., Marino, J., Amanat, F., Krammer, F., Hahn-Holbrook, J., Zolla-Pazner, S., Powell, R. L. Evidence of a significant secretory-IgA-dominant SARS-CoV-2 immune response in human milk following recovery from COVID-19. *MedRxiv* 2020; 5.
12. Campanhon, I. B., da Silva, M. R. S., de Magalhães, M. T. Q., Zingali, R. B., Bezerra, F. F., Torres, A. G. Protective factors in mature human milk: a look into the proteome and peptidome of adolescent mothers' breast milk. *Br J Nutr* 2019; 122 (12): 1377-1385.
13. Zhang, Q., Cundiff, J. K., Maria, S. D., McMahon, R. J., Wickham, M. S., Faulks, R. M et al. Differential digestion of human milk proteins in a simulated stomach model. *J Proteome Res* 2013; 13 (2): 1055-1064.
14. Dall'Asta, C., Florio, P., Lammardo, A. M., Prandi, B., Mazzeo, T., Budelli, A et al. Development of an in vitro digestive model for studying the peptide profile of breast milk. *Int. J Food Sci* 2015 66 (4): 409-415.
15. Wan, J., Cui, X.W., Zhang, J., Fu, Z.Y., Guo, X.R., Sun. L.Z et al. Peptidome analysis of human skim milk in term and preterm milk. *Biochem Biophys Res Commun* 2013; 438 (1)2: 36-41.
16. Dallas, D.C., Smink, C.J., Robinson, R.C., Tian, T., Guerrero, A., Parker, E.A et al. Endogenous human milk peptide release is greater after preterm birth than term birth. *J Nutr* 2015; 145 (3): 425-33.
17. Lowry, Oliver H. Protein measurement with the Folin phenol reagent. *J Biol Chem* 1951, 193: 265-275.
18. Keller, R. P., Neville, M. C. Determination of total protein in human milk: comparison of methods. *Clin Chem* 1986; 32: 120-123.

19. Nielsen, S.D., Beverly, R.L., Qu, Y., Dallas, D.C. Milk bioactive peptide database: A comprehensive database of milk protein-derived bioactive peptides and novel visualization. *Food Chem* 2017; 232: 673-682.
20. Torrent, M., Di Tommaso, P., Pulido, D., Nogués, M.V., Notredame C et al. AMPA: an automated web server for prediction of protein antimicrobial regions. *Bioinformatics* 2012, 28 (1): 130-1.
21. Joseph, S., Karnik, S., Nilawe, P., Jayaraman, V. K., Idicula-Thomas, S. ClassAMP: A prediction tool for classification of antimicrobial peptides. *IEEE/ACM Trans Comput Biol Bioinform* 2012; 9:1535–1538.
22. Wang, G., Li,X., Wang, Z. APD3: the antimicrobial peptide database as a tool for research and education. *Nucleic Acids Res* 2015; 44 (D1): D1087-D1093.
23. Théolier, J., Jérémie, T., Ismail, F., Julie, J., Riadh, H. MilkAMP: a comprehensive database of antimicrobial peptides of dairy origin. *Dairy Sci Technol* 2013; 94 (2): 181–193.
24. Kumar, R., Chaudhary, K., Sharma, M., Nagpal, G., Chauhan, J. S Singh, S et al. AHTPDB: a comprehensive platform for analysis and presentation of antihypertensive peptides. *Nucleic Acids Res* 2015; 43 (D1): D956-D962.
25. Minkiewicz, P., Dziuba, J., Iwaniak, A., Dziuba, M., Darewicz, M. BIOPEP database and other programs for processing bioactive peptide sequences. *J AOAC Int* 2008; 91(4): 965-980.
26. DeLano, W. L., Bromberg, S. *PyMOL user's guide*. DeLano Scientific LLC 2004; 629.
27. Maupetit, J., Derreumaux, P., Tuffery, P. PEP-FOLD: an online resource for de novo peptide structure prediction. *Nucleic Acids Res* 2009; 37: W498-W503.
28. Madureira, A. R., Pereira, A., Castro, P. M.,Pintado, M. Production of antimicrobial chitosan nanoparticles against food pathogens. *J Food Eng* 2015; 167: 210-216.
29. Dong, Y., Chi, X., Hai, H., Sun, L., Zhang, M., Xie, W.F et al. Antibodies in the breast milk of a maternal woman with COVID-19. *Emerg Microbes Infect* 2020; 9: 1467–1469.
30. Van der Weerden, N. L., Bleackley, M. R., Anderson, M. A. Properties and mechanisms of action of naturally occurring antifungal peptides. *Cell Mol Life Sci* 2013; 70 (19): 3545-3570.



31. Sharma, A., Singh, G., Bhatti, J. S., Gill, S. K., Arya, S. K. Antifungal peptides: Therapeutic potential and challenges before their commercial success. *Int J Biol Macromol* 2024; 137957.
32. Bersani, I.; Piersigilli, F.; Goffredo, B. M.; Santisi, A.; Cairoli, S.; Ronchetti, M et al. Antifungal drugs for invasive candida infections (ICI) in neonates: Future perspectives. *Front Pediatr* 2019; 7: 375.
33. Deslouches, B., Di, Y. P. Antimicrobial peptides with selective antitumor mechanisms: prospect for anticancer applications. *Oncotarget* 2017; 8 (28): 46635–46651.
34. Zhang, R., Xu, L., Dong, C. Antimicrobial Peptides: An Overview of their Structure, Function and Mechanism of Action. *Protein Pept Lett* 2022; 29(8):641-650.
35. Porcelli, F., Ramamoorthy, A., Barany, G., Veglia, G. On the role of NMR spectroscopy for characterization of antimicrobial peptides. *Membrane Proteins* 2013: 159-180.
36. Bortolus, M., Dalzini, A., Toniolo, C., Hahm, K. S., Maniero, A. L. Interaction of hydrophobic and amphipathic antimicrobial peptides with lipid bicelles. *J Pept Sci* 2014; 20: 517–525.
37. Gamage, D. G., Gunaratne, A., Periyannan, G. R., Russell, T. G. Applicability of Instability Index for In vitro Protein Stability Prediction. *Protein Pept Lett* 2019; 26(5):339-347.
38. Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M. R., Appel, R. D., Bairoch, A. *Protein identification and analysis tools on the ExPASy server*. Humana press 2005; 571-607.
39. Ovali, F., Ciftci, I. H., Çetinkaya, Z. A. F. E. R., Bükülmez, A. Y. Ş. E. G. Ü. L. Effects of human milk fortifier on the antimicrobial properties of human milk. *J perinatol* 2006; 26(12): 761-763.



## Figures' captions

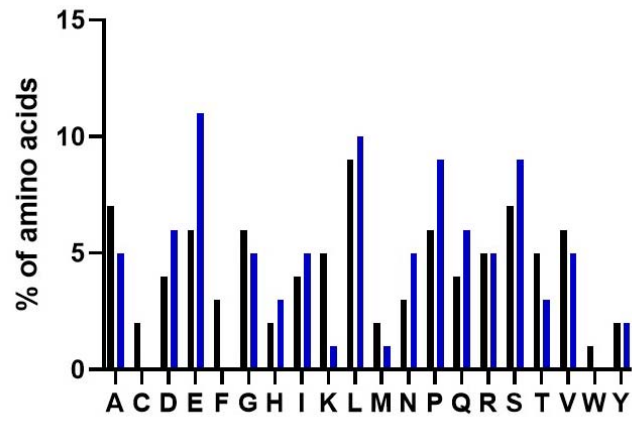
**Fig. 1.** Frequency of amino acids found in the endogenous peptides' sequences in human colostrum (■) annotated by mass spectrometry (nLC-HRMS/MS Orbitrap) compared to the frequency of amino acids present in human proteins (■) calculated using the Genscript program (<https://www.genscript.com/tools/codon-frequency-table>). The one-letter coding of amino acid names was used in this Figure.

**Fig. 2.** Antimicrobial activity of endogenous peptides identified by mass spectrometry (nLC-HRMS/MS Orbitrap) in human colostrum using the ClassAMP program. Antifungal activity (45%), antibacterial activity (38%), and antiviral activity (17%).

**Fig. 3.** Structures of endogenous peptides in human colostrum predicted by PEPFOLD3 program. Color-coding of peptides' secondary structure: pink,  $\alpha$ -helix; yellow,  $\beta$ -sheet; grey, random structure.

**Fig. 4.** Ribbon representation of the 3D structures of three proteins identified as precursors of endogenous peptides in human colostrum. The predicted 3D structures were extracted from PDB and visualized using PyMOL software. (A) P02788, lactoferrin (PDB ID: 1H43), with the peptide AGLAPYKLRPV shown in orange; (B) P01591, immunoglobulin J (PDB ID: 6LX3), and in pink the peptide DPNEDIVERNI; (C) P01833, polymeric immunoglobulin receptor protein (PDB ID: pdb2ocw), with the peptide IRLDIQGTGQLLF shown in blue.

**Figure 1**



**Figure 2**

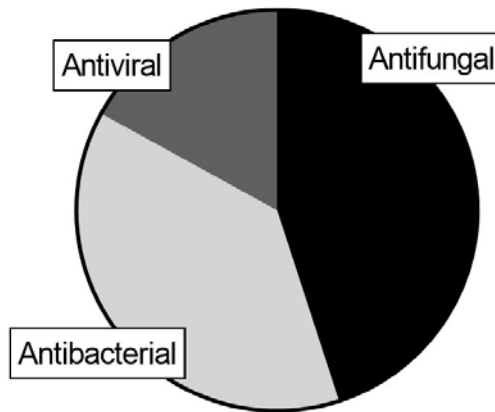
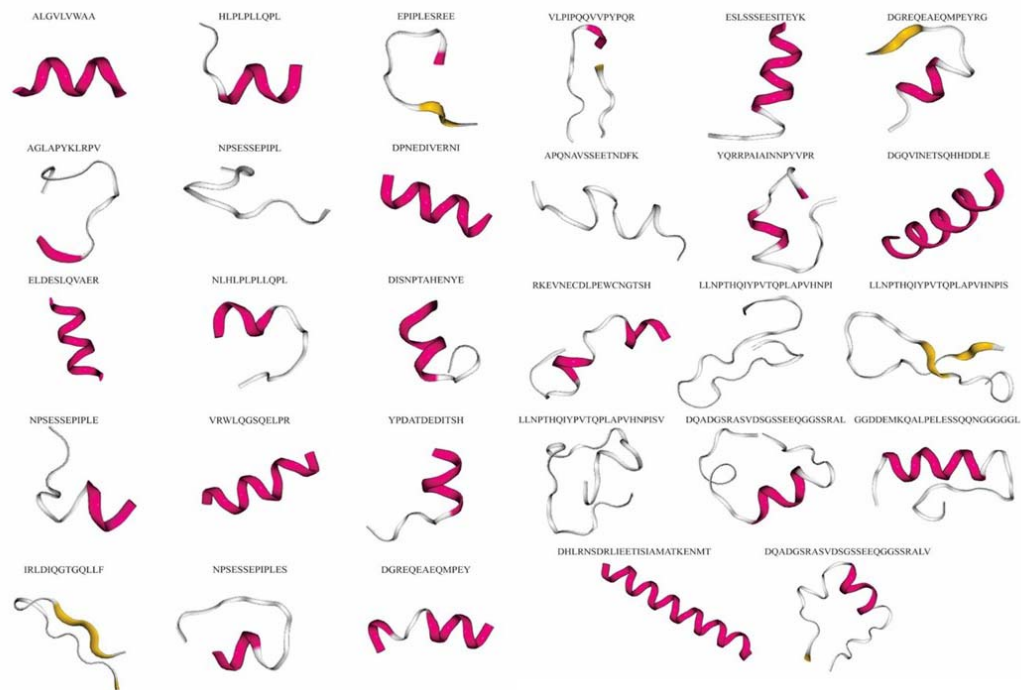
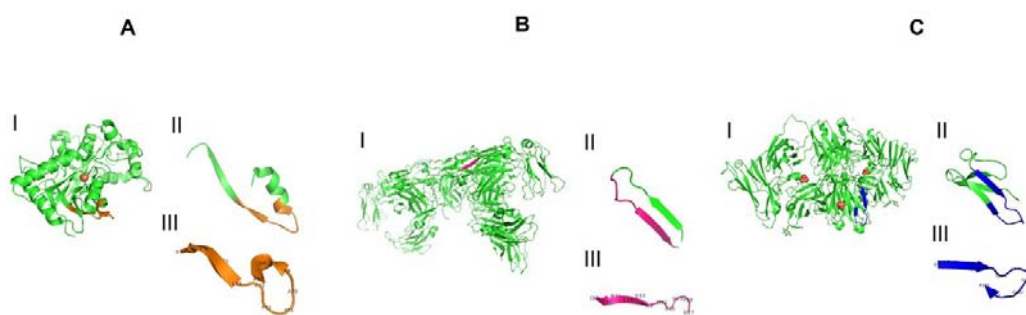


Figure 3



**Figure 4**



**Table 1:** Peptide sequences in the endogenous fraction (<10 kDa) of human colostrum (pooled sample, from  $n = 24$  adult donors) annotated by mass spectrometry (nLC-HRMS/MS, Orbitrap), and their precursor proteins.

Peptide sequence	Exact mass [Da]	Precursor protein accessions	Precursor protein name
DQADGSRASVDSGSSEEQGGSSRAL	2453.0776		Polymeric immunoglobulin receptor
IRLDIQGTGQLLF	1473.8429	P01833	
DQADGSRASVDSGSSEEQGGSSRALV	2552.1460		
ALGVLVWAA	899.5354	O60831	PRA1 family protein 2
HLPLPLLQPL	1140.7144		
LLNPTHQIYPVTQPLAPVHNPIV	2648.4564		
LLNPTHQIYPVTQPLAPVHNPI	2462.3560		
LLNPTHQIYPVTQPLAPVHNPI	2549.3880	P05814	Beta-casein
VLPQPQVVPYPQR	1633.9430		
ESLSSEESITEYK	1588.7229		
NLHLPLPLLQPL	1367.8414		
NPSESSEPIPL	1169.5690		
DISNPTAHENYE	1389.5922		
NPSESSEPIPLE	1298.6116	P47710	Alpha-S1-casein
NPSESSEPIPLES	1385.6436		
EPIPLESREE	1198.5955		
VRWLQGSQELPR	1468.8024	P0DOX2	Immunoglobulin alpha-2 heavy chain
DPNEDIVERNI	1313.6337	P01591	Immunoglobulin J chain
YQRRPAIAINNPYVPR	1928.0618	P07498	Kappa-casein
AGLAPYKLRPV	1184.7150	P02788	Lactotransferrin
GGDEMCKALPELESSQQNGGGGGL	2474.1105	Q9UK61	Protein TASOR
ELDESLQVAER	1288.6384	P10909-2	Isoform 2 of clusterin
DGREQEAEQMPEY	1581.6491		
DGREQEAEQMPEYRG	1794.7717	Q13410	Butyrophilin subfamily 1 member A1
DHLRNSDRLIETISIAMATKENMT	2888.4245		
APQNAVSSSEETNDFK	1636.7454	O95249	Golgi SNAP receptor complex member 1
YPDATDEDITSH	1363.5653	P10451	Osteopontin
RKEVNECDLPEWCNGTSH	2116.9180	Q9UKF5	Disintegrin and metalloproteinase domain-containing protein 29
DGQVINETSQHDDLE	1836.8000	P08670	Vimentin

**Table 2.** Physicochemical characteristics of endogenous peptides (fraction < 10 kDa) in human colostrum annotated by mass spectrometry (nLC-HRMS/MS Orbitrap) evaluated by ProtParam software.

Sequence of amino acids	pI	MW (Da)	Instability index	Aliphatic index	GRAVY index*
ALGVLVWAA	5.57	899.1	-7.81*	184.4	2.233
HLPLPLLQPL	6.74	1140.4	99.38	195	0.75
EPIPLESREE	4.09	1198.3	96.5	78	-1.42
AGLAPYKLRPV	9.99	1184.4	71.41	115.5	0.191
NPSESSEPIPL	3.79	1169.2	108.91	70.91	-0.855
DPNEDIVERNI	3.92	1313.3	70.05	97.27	-1.264
ELDESLQVAER	4	1288.3	49.38	106.36	-0.836
NLHLPLPLLQPL	6.74	1367.7	84.48	195	0.65
DISNPTAHENYE	4.13	1389.4	-6.59*	40.83	-1.567
NPSESSEPIPLE	3.67	1298.3	100.67	65	-1.075
VRWLQGSQELPR	9.57	1468.6	136.22	89.17	-0.95
YPDATDEDITSH	3.84	1363.3	29.22*	40.83	-1.333
IRLDIQGTGQLLF	5.84	1473.7	-3.83*	150	0.515
NPSESSEPIPLES	3.67	1385.4	108.51	60	-1.054
DGREQEAEQMPEY	3.91	1581.6	100.85	7.69	-2.2
VLPIPQVVVPYQQR	8.72	1633.9	79.44	117.86	-0.129
ESLSSSEESITEYK	4.09	1588.6	115.11	55.71	-1.114
DGREQEAEQMPEYRG	4.25	1794.8	71.8	6.67	-2.233
APQNAVSEETNDFK	4.14	1636.6	57.36	32.67	-1.213
YQRRPAIAINNPYVPR	10.9	1928.2	74.78	79.38	-0.912
DGQVINETSQHDDLE	4.1	1836.8	16.7*	66.88	-1.488
RKEVNECDLPEWCNGTSH	4.83	2117.3	43.49	37.78	-1.333
LLNPTHQIYPVTQPLAPVHNPI	6.92	2462.8	51.03	119.55	-0.023
LLNPTHQIYPVTQPLAPVHNPIIS	6.92	2549.9	49.25	114.35	-0.057
LLNPTHQIYPVTQPLAPVHNPIISV	6.92	2649	47.61	121.7	0.121
DQADGSRASVDSGSSEEQGGSSRAL	4.11	2453.4	72.1	39.2	-1.092
GGDDEMKGALPELESSQNGGGGGL	3.83	2474.5	83.21	50.8	-1.052
DHLRNSDRLIEETISIAMATKENMT	4.9	2889.2	53.62	86	-0.632
DQADGSRASVDSGSSEEQGGSSRALV	4.11	2552.5	69.71	48.85	-0.888

\*Grand average of hydropathicity