



OPEN

Peptide profiling in cow urine reveals molecular signature of physiology-driven pathways and in-silico predicted bioactive properties

Rohit Kumar, Syed Azmal Ali, Sumit Kumar Singh, Vanya Bhushan, Jai Kumar Kaushik, Ashok Kumar Mohanty & Sudarshan Kumar✉

Peptidomics allows the identification of peptides that are derived from proteins. Urinary peptidomics has revolutionized the field of diagnostics as the samples represent complete systemic changes happening in the body. Moreover, it can be collected in a non-invasive manner. We profiled the peptides in urine collected from different physiological states (heifer, pregnancy, and lactation) of Sahiwal cows. Endogenous peptides were extracted from 30 individual cows belonging to three groups, each group comprising of ten animals (biological replicates $n = 10$). Nano Liquid chromatography Mass spectrometry (nLC-MS/MS) experiments revealed 5239, 4774, and 5466 peptides in the heifer, pregnant and lactating animals respectively. Urinary peptides of <10 kDa size were considered for the study. Peptides were extracted by 10 kDa MWCO filter. Sequences were identified by scanning the MS spectra ranging from 200 to 2200 m/z . The peptides exhibited diversity in sequences across different physiological states and in-silico experiments were conducted to classify the bioactive peptides into anti-microbial, anti-inflammatory, anti-hypertensive, and anti-cancerous groups. We have validated the antimicrobial effect of urinary peptides on *Staphylococcus aureus* and *Escherichia coli* under an in-vitro experimental set up. The origin of these peptides was traced back to certain proteases viz. MMPs, KLKs, CASPs, ADAMs etc. which were found responsible for the physiology-specific peptide signature of urine. Proteins involved in extracellular matrix structural constituent (GO:0005201) were found significant during pregnancy and lactation in which tissue remodeling is extensive. Collagen trimers were prominent molecules under cellular component category during lactation. Homophilic cell adhesion was found to be an important biological process involved in embryo attachment during pregnancy. The in-silico study also highlighted the enrichment of progenitor proteins on specific chromosomes and their relative expression in context to specific physiology. The urinary peptides, precursor proteins, and proteases identified in the study offers a base line information in healthy cows which can be utilized in biomarker discovery research for several pathophysiological studies.

Excretory biological fluids such as urine, saliva, milk, tear, mucus and sweat are significantly important for the maintenance of homeostasis in normal physiological conditions. Urine being a glomerular filtrate of blood is capable of summarizing the events that occur in the body as a result of changing physiology or pathological conditions. The systemic changes are well reflected by qualitative and quantitative alterations in the urine composition. No wonder, urine has been considered as excellent sample for the discovery and detection of biomarkers associated with general health and disease. Endogenous peptides and proteins secreted in urine have proven as hallmarks of various pathophysiological changes and have emerged as a better option over other biological fluids as it can be obtained in a large volume without much maneuver in dairy animals.

The quantitative estimation of urinary biomarkers is affected by changes in the volume of voided urine¹. In the absence of a standard baseline profile of peptides in the urine, identification and quantitation of biomarkers

ICAR-National Dairy Research Institute, Cell Biology and Proteomics Lab, Animal Biotechnology Center (ABTC), Karnal, Haryana 132001, India. ✉email: kumarsudershan@gmail.com

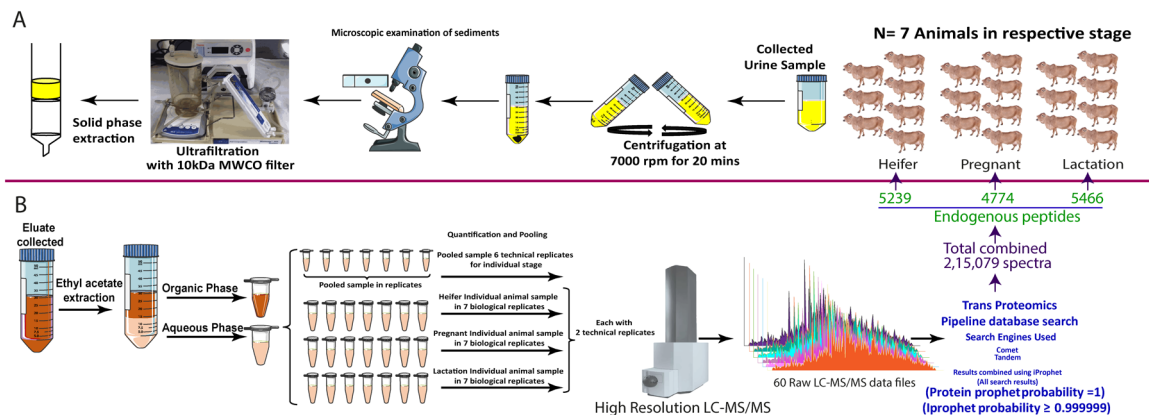


Figure 1. (A) Flow diagram showing collection and processing of samples from three different physiological states of Sahiwal cow followed by separation of small-sized peptides (<10 kDa) by ultrafiltration assembly and its extraction by C18 solid-phase extraction method. (B) The solid phase extraction (SPE) method. The bound analytes in the column were first eluted in 60% ACN, 0.1% TFA followed by the extraction of peptides in the aqueous phase using ethyl acetate. An equal amount of peptides from each sample (biological replicates) was pooled group-wise and quantified, for nLC-MS/MS analysis in triplicate (technical replicates).

usually becomes challenging. The peptides present in urine remain stable for a long time. The process of degradation of urinary proteins into resultant peptides is completed by the time of sample collection, whereas protease activation in fluids like the serum, plasma, and saliva remains continued even after sample collection that affects the actual result^{2,3}.

Besides, the metabolites, minerals, and salts in the glomerular filtrate, urine is composed of peptides and proteins originating from tubular secretion, secreted exosomes, and epithelial cells shed from the kidney and urinary tract⁴. Urine carries more local information as indicated from studies that showed that 70% of the proteome comprises secretions down from the kidney to urinary tracts and the remaining 30% originates from plasma⁵. Under the Human Kidney and Urine Proteome Project, comprehensive studies has been carried out to find novel biomarkers associated with kidney diseases, and to unravel the proteome of urine in normal and diseased conditions^{6–9}. Extensive investigation of human urinary endogenous peptides has been carried out in the context of various diseases stating their strong clinical relevance^{10–12}. However, studies in bovine are still in the early phase. Around 1550 proteins have been identified in the urine of healthy Karan fries cows^{13,14}. Although reports related to protein-based biomarkers in saliva and urine are present^{14–16}, the peptidome based studies are very scanty in bovine and most of the studies belong to milk. Milk has been explored for its peptide and protein profiling, identifying around 8559 unique peptides and 6210 proteins, respectively^{17,18}. Lack of studies in bovine urinary peptidome makes it difficult for researchers to explore and investigate urine for its application in diagnosis and clinical application. A comprehensive urinary peptidome database encompassing diverse peptides tracing back to systemic and local origin which are processed by a variety of endogenous proteases can tell a defined story of animal physiology which can be referred to in the future for studies involving bioactive peptides, and biomarker discovery for clinical diagnosis and prognosis.

Although contemporary studies are inclined towards the discovery of biomarkers associated with a particular condition or disease, the data generated by MS/MS analysis of urine can track the changes in the proteome & peptidome not only in a condition-specific manner but also in establishing physiology-specific bio-molecular signatures present in it. In that direction, the foremost thrust in the field of clinical proteomics is the profiling of such endogenous peptides.

The current study reveals urinary peptidome in the Sahiwal breed of cows across different physiological stages and their resourcefulness for various applications. Urine samples were collected from 10 healthy animals in each group, namely heifer, pregnant and lactating. The peptides were identified and characterized to signify its importance in deriving meaningful information in the context of physiological changes and several other attributes associated with cow urine.

Results

Endogenous peptide profile in cow urine. The endogenous peptides were purified using a solid-phase ethyl acetate extraction method (Fig. 1A). The total collection of physiologically distinct (heifer, pregnant, and lactation) urine samples resulted in 2,15,079 spectra (Fig. 1B). The pooled samples (n = 10) in each physiological state resulted in the identification of 5239, 4774, and 5466 endogenous peptides in the heifer, pregnant, and lactation, respectively. The dataset was analyzed using Trans Proteomics Pipeline (TPP) with three independent search engines. The complete results summary has been provided in Table 1. The protein prophet and iProphet value of 1 and 0.9999 respectively, along with less than 1% FDR was used as a cutoff to select the highly confident peptides (Supplementary Table 1) (Fig. 1B). Initial screening of mapped peptides showed the highly diversified nature of the endogenous peptidome.

Total raw files	Conditions	Unassigned total spectra#	Sum total spectra (after assignment)	Endogenous De novo Peptides	Proteins (Protein prophet probability=1) (Iprophet probability \geq 0.999999)
20	Heifer	214471	27083	5239	2092 (900)*
20	Pregnant	218556	27336	4774	2002 (755)*
20	Lactation	212208	27025	5466	1946 (860)*

Table 1. # Total number of unassigned spectra acquired in all the raw data mapped in COMET search engine using no enzyme search parameter. *Proteins reported with bold letters asterisk symbol are identified with more than 10 percent sequence coverage peptides of respective protein and all the peptides identified through minimum of 2 independent spectra.

Molecular characteristics of urinary peptides. The frequency distribution of molecular weight, peptide length, and amino acid composition of peptides belonging to three different groups have been shown in (Fig. 2). The peptide from different groups were found to have a molecular size of less than 10 kDa (Fig. 2A). In heifer and lactating animals' urine, low molecular weight peptides ranging from 1.4–1.5 kDa were more prevalent. In contrast, in pregnant animals' urine, peptides of relatively large size in the range of 1.8, 2.2, and 2.9 kDa were more prevalent (Fig. 2A). The peptides length distribution comparison showed equal frequency in three conditions (Fig. 2B). However, amino acid composition indicated that alanine, glycine, leucine, proline, and serine (in decreasing order of abundance) were the most frequently occurring amino acids (Fig. 2C) which reflected a similar pattern across the three physiological states under study. The band pattern of peptides in Tricine-gel showed similar results across different groups (Fig. 3A). The present study focused on endogenous peptides by filtering out the molecules larger than 10 kDa using a 10 kDa cut-off membrane. The observation agreed with the findings of molecular characteristics (Fig. 2) that the urinary peptides are of low molecular weight, i.e., less than 10 kDa. In summary, the results suggest that endogenous peptides in urine are highly diversified in sequences but contain comparable molecular characteristics. It guided us to identify the cause and origin and impact of underlying diversity in the peptidome.

Functional attributes associated with urinary peptidome. Urine is full of secretions which offer protection to the urogenital tract from the environmental pathogens. We characterized urinary peptides with several biological activities using in-silico and in-vitro experiments. In-silico experiments were conducted to classify the bioactive peptides into anti-microbial, anti-inflammatory, anti-hypertensive, and anti-cancerous groups. The notion was to identify the mass spectrometer-based profiled cow urinary peptides in all the classes and to narrow down the list of peptides to a few significant candidates which exhibited wide spectrum of activities. Excitingly, we identified 15 peptides displaying in-silico activity of all types of processes (Fig. 3B). Anti-cancer and antihypertensive sets shared 211 peptides constituting 6.3%. The antimicrobial (n = 551) and anti-inflammatory (n = 607) peptides shared 36 sequences (Fig. 3B). Sequence analysis of all the peptides showed that glycine amino acid was a common and most abundant residue in all the classified peptides (Fig. 3C, D, E, and F). This bears an important message in the designing and synthesis of broad-spectrum bioactive peptides that can perform multiple functions.

A total of 551 high scoring antimicrobial peptide sequences were used to determine the consensus motif of amino acids. The average hydrophobicity in the sequences was -0.13 with an average charge of $+1.8$, indicating that sequences are hydrophilic and cationic. The sequence logo showed the dominant presence of positively charged residues like Arg and Lys, whereas Gly and Leu were dominant hydrophobic residues (Fig. 3C). A total of 607 sequences were predicted to possess anti-inflammatory activity. A positional conservation study showed that leucine residue was relatively dominant at the first 10th position from the N-terminal end in all anti-inflammatory peptides. Positional conservation study at N-terminal of anti-inflammatory peptides showed Leu, Tyr, Ser, Arg, and Glu to be highly conserved residues. While its position may vary, Leu residue seems to be central to the anti-inflammatory activity (Fig. 3D). We identified 1767 sequences predicted with antihypertensive activity with Pro as dominant residue at all positions (Fig. 3E). In comparison, 724 unique sequences were predicted to possess the activity mentioned above (Fig. 3F) (Supplementary Table 2). For validation of the in-silico finding, we tested the antimicrobial activity of isolated peptides against two pathogens, *E. coli* and *S. aureus* (Fig. 3G and H). The pure peptide extract showed a significant zone of inhibition in the disk diffusion method with a mean inhibition zone of 1.22 ± 0.11 (SD) cm; and SEM: ± 0.03 cm on *E. coli* and mean inhibition zone of 1.22 ± 0.10 (SD) cm; SEM: ± 0.03 cm on *S. aureus*. The results of in-vitro experiments suggest that urinary peptides possess strong antimicrobial activity.

Bovine urine degradome. Manual curation of 22 selected proteases resulted in the discovery of an average of 7215 protease sites (Supplementary Table 3). The urinary peptides were investigated for four residues either from N or C terminal and peptides were sorted based on possible protease that might be involved in its release from the precursor protein. We found physiology-driven variation in the number of sequences derived from proteases. We further validated our findings using the web-based tool: Proteasix. It determines confidence thresholds of predicted proteolysis by using MEROPS specificity weight matrices for experimentally confirmed cleavages. MEROPS is a database that provides information about protease substrate sequence specificity in terms of the specificity weight matrix. Based on experimentally confirmed cleavages, the matrix shows the fre-

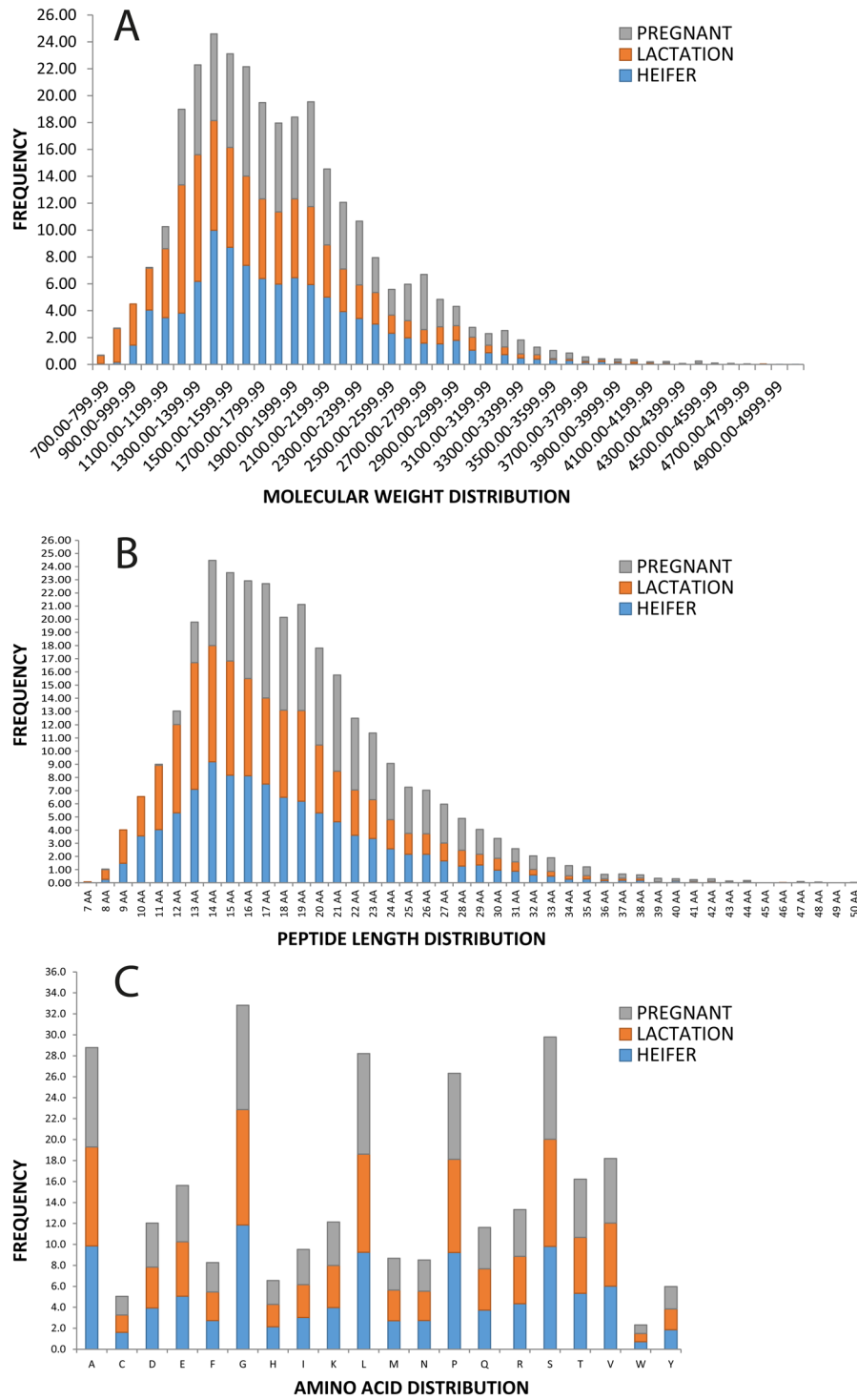


Figure 2. (A) Distribution of molecular weight of peptides along with their relative abundance. (B) Peptide length distribution and (C) Distribution of amino acid residues in three different groups.

quency of amino acids at every position of the site. The output data were refined by removing the cleavages that were having specificity below 80%. We determined the common protease activity in all three physiological conditions and found that 54 proteases out of 62 potential proteases (85.7%) were common (Fig. 4A). No protease enzyme could be uniquely associated with pregnancy. However, two unique proteases were reported in the heifer and lactating groups (Fig. 4B). Pearson correlation-based group-wise comparison (heifer, pregnant, and lactation) is 0.99 among each other, describing highly correlated data. The heat map represents the total number of

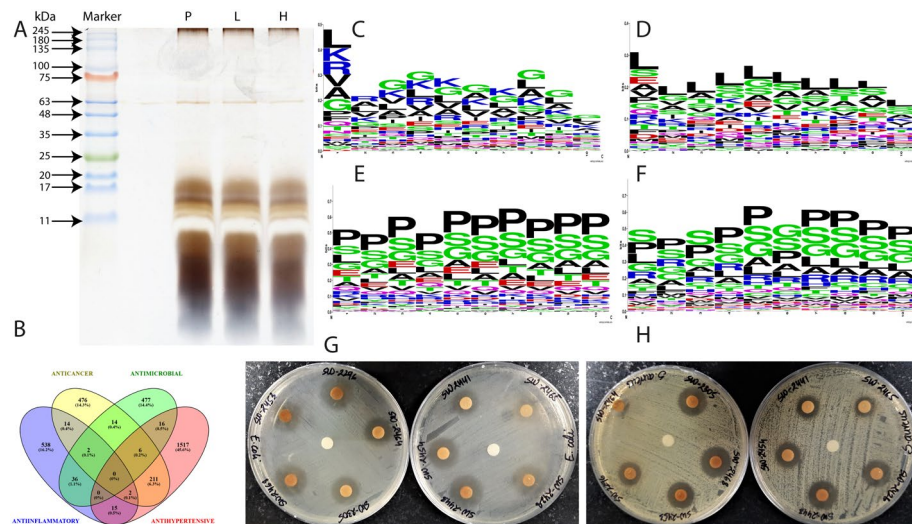


Figure 3. (A) After subjecting <10 kDa cow urine fraction to Solid Phase extraction, the eluate was collected in 60% ACN, 0.1% TFA. Ethyl acetate extraction of the eluate was carried out and the aqueous phase was separated and dried in Speed Vac. Aqueous extract pooled from different physiological states viz. Heifer, Lactation, and Pregnant (left to right) of Sahiwal cattle were then visualized by tricine-SDS PAGE. (B) Venn diagram showing the number of common peptide sequences shared by different classes of predicted bioactivity. (C) N terminal (10 amino acids) sequence logos were generated for the sequences predicted with certain bioactivity. Antimicrobial Peptide sequences showed glycine, lysine, and leucine as dominant residues. (D) Sequence logo for predicted Anti-Inflammatory sequences, every position is predominated by leucine residue. (E) Sequence logo for predicted Antihypertensive sequences with Proline as dominant residue. (F) Sequence logo for predicted Anticancer sequences with Proline as dominant residue. Both antihypertensive and anticancer sequences show proline as the dominant amino acid at almost every position. (G) 20 μ L of Solid Phase Extract eluate from 10 individual animals (Heifer) was coated on 6mm sterile disc and placed on *Staphylococcus aureus* lawn (equivalent to 0.5 McFarland units). Zone of Inhibition was observed after 12–18 h of incubation. Disc in the center is a negative control containing 20 μ L of BSA digest. (H) Peptide extracts activity against *Escherichia coli* (equivalent to 0.5 Mc Farland). Disc in the center is a negative control containing 20 μ L of BSA digest.

target peptides identified in the data to specific protease (Fig. 4C) and similarity Pearson comparison (Fig. 4D). The occurrence of Matrix Metalloproteases (MMPs) isoforms was wide spread across all the physiological states which were mapped to their target proteins (Fig. 4E). Several isoforms of KLK enzymes were identified with potential cleavage sites on the target proteins leading to the formation of bovine urinary degradome (Fig. 4F). Protease wise comparison and measure of Pearson correlation (Fig. 4D) identified eight different protease enzyme (CASP2, CASP8, FURIN, LGMN, MMP10, MMP17, PCSK1, and PRSS3) that were in disagreement with physiological state and uncorrelated with one another (Fig. 4G). The profile plot and heatmap representing the number of protease targets in all three stages revealed that enzymatic degradation of target proteins during pregnancy is somewhat slow and suppressed (Fig. 4F and A).

Chromosomal mapping of stage-specific bovine endogenous urinary peptidome. Total spectra and protein information identified from the search engine using the UniProt database were represented as the interconnecting bars in the circle (Fig. 5A). Notable results are the identification of the denovo peptides mapped to all the 29 + X chromosome. Interesting facts uncovered was the uneven distribution of the protein expression from the chromosomes in a physiological stage-specific manner. The expression of proteins from the chromosomes was drastically inconsistent at different physiological states (Heifer, Lactation, and Pregnant). Bubble plot-based heat map analysis showed the total protein counts represented from respective individual chromosomes (Fig. 5A, left side) and percentage contribution of the proteins in terms of enrichment from the total number of proteins to the coded in the bovine genome (Fig. 5A, Right side).

Chromosomes 3, 5, 7, 11, 18, and 19 were responsible for hosting the maximum number of endogenous peptides on the genome, whereas chromosomes 12, 20, 24, 27, and 28 gave rise to a minimum number of peptides that are identifiable in the urine. The unbiased analysis of enrichment percentage for all the chromosomes mapped sequences showed the high enrichment of respective peptides from chromosome 14, 20, 21, 22, 25, and 28 in the stages of heifer and lactation. Interestingly, chromosome 12 was rich in the release of peptides in pregnant animal urine samples, whereas all other chromosomes contributed equally. The same results were obtained from the distance-based correlation analysis among chromosomes. We found the distinct expression values of chromosome 8, 12, and 15 in comparison to others (Fig. 5B). Our finding raises a question: what are the reasons responsible for the high expression from certain chromosomes in heifer and lactation stages while the low contribution of coded peptides and proteins from the same chromosomes in the pregnant state? Hitherto, factors that decide the differential expression of chromosomes in different physiological states is not known.

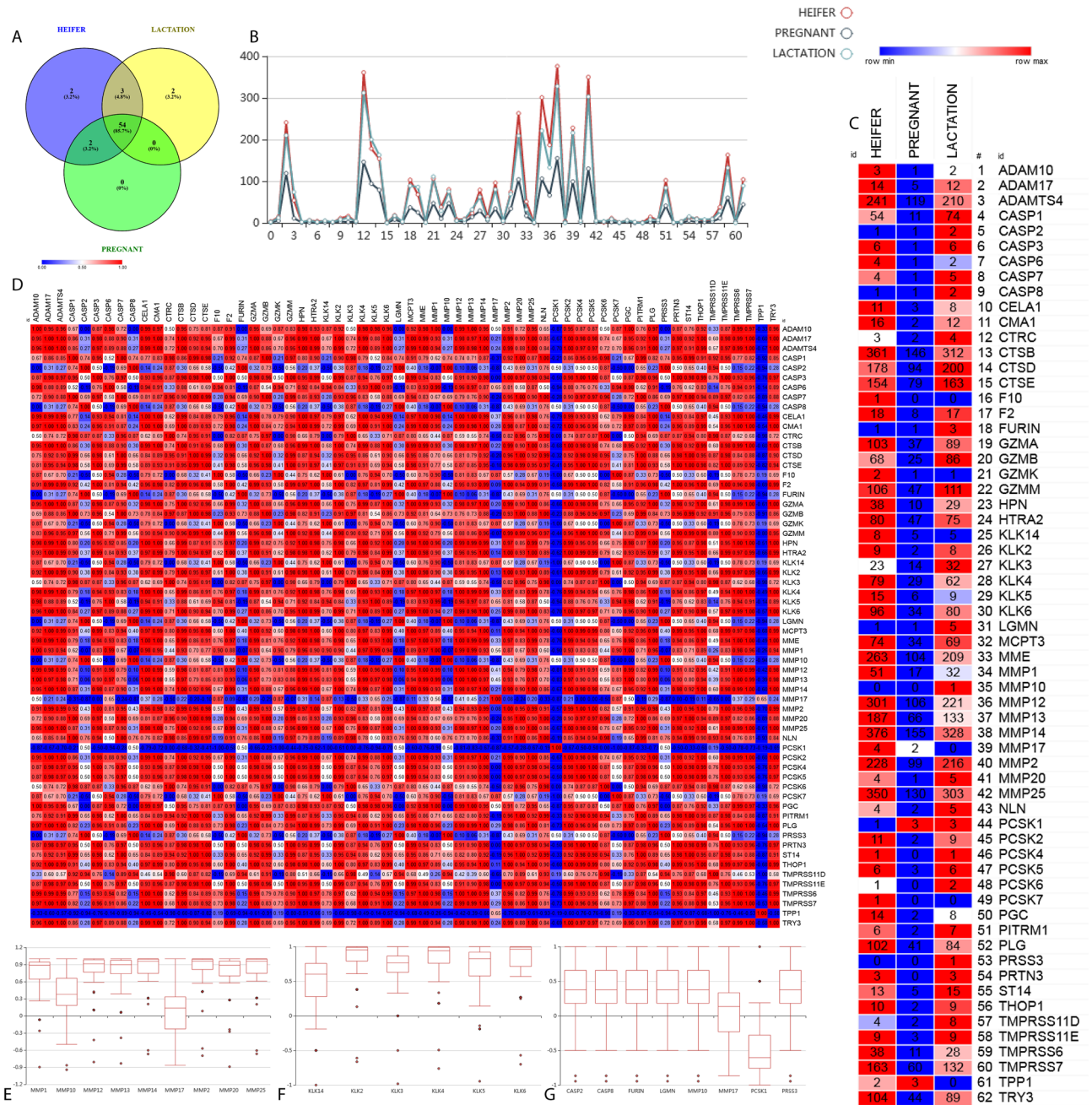


Figure 4. Degradome Analysis: (A) Venn Diagram describing the common and unique enzyme identified among the stages in the protein data mapping. (B) Profile plot depicting the quantitative values of target proteins (Y-axis) map to 62 different protease enzymes (X-axis). (C) Quantitative heat map presenting the protease enzyme class with the respective target number in three conditions. (D) Pearson correlation-based similarity matrix plot of all protease enzymes and quantitative target relationship. (E) Distribution of MMP isoforms identified in protein mapping and respective target determined in peptidome. (F) Distribution of isoforms of KLK enzymes isoforms identified in protein mapping and respective target determined in peptidome. (G) Protease wise comparison for Pearson correlation of 8 different enzyme (CASP2, CASP8, FURIN, LGMN, MMP10, MMP17, PCSK1, and PRSS3) in disagreement with physiological state and uncorrelated with one another.

We suggest that physiology specific expression of proteins is probably the reason which dictates the profile of peptides in urine. Nonetheless, further studies on genome-wide proteome analysis are required to answer these questions. The heatmap analysis for 483 common parental proteins mapped in all three stages with significant St Peter's values showed the clustering pattern among the proteins as defined by the Kmean clusters (Fig. 5C). Keeping the heifer stage as the control, we determined the fold change difference among the stages. The results showed the identification of five proteins which are more than 9 fold up-regulated in pregnant and lactation

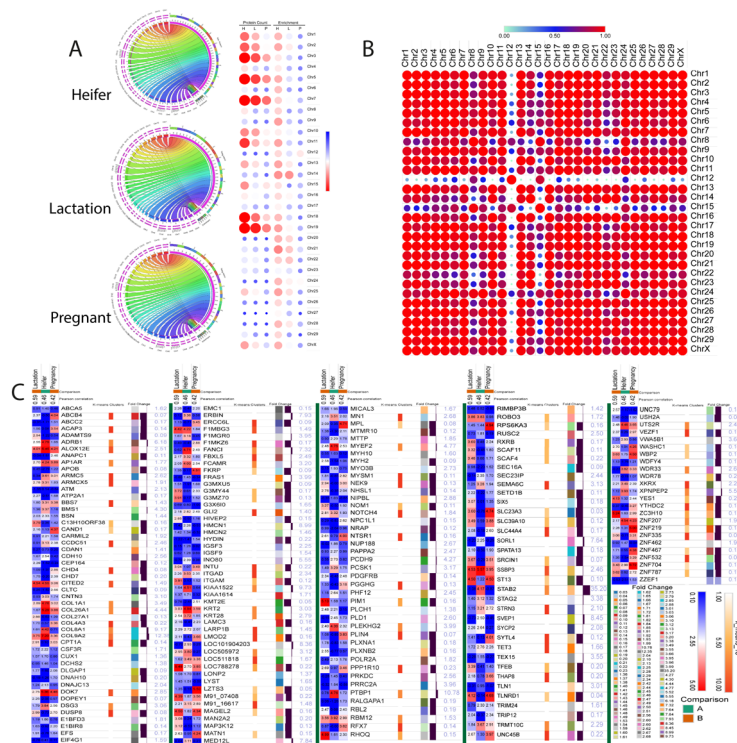


Figure 5. Peptide to protein mapping: (A) Circos plots describing the mapping of peptides to specific chromosomes through parent mapping. The right-hand side details the information of protein count heat map and enrichment of in three stages. (B) Pearson correlation-based similarity matrix plot of all chromosomes and associated target proteins determined with respect to endogenous peptides. (C) Heat map presentation of proteins quantitative data for common parental proteins among three stages. Fold change is calculated using the heifer as the control for lactation and pregnancy. Proteins were grouped based on the k-mean clustering and fold change calculate with respective color coding.

stages, the proteins are STAB2 (FC: 35.20), KIAA1522 (FC: 9.73), COL8A1 (FC: 9.17), COL9A2 (FC: 12.35), and PTBP1 (FC: 10.78). It shows the importance of these five proteins in pregnancy and lactation.

Tracing back Peptidome to physiology specific functional annotation. We compared all of the protein concentrations identified through StPeter among the physiological states (Fig. 6A). The analysis identified the 525, 465, 458 common proteins between heifer versus lactation, heifer versus pregnant, and pregnant versus lactation, respectively. The total numbers of average 1500 proteins were identified uniquely in all three conditions. The molecular function analysis showed the identification of parental proteins contributing to the multiple cell replication process (Fig. 6B) with respective p values reported (Supplementary Table 4). The processes are RNA polymerase DNA binding, ATP binding, promoter-specific chromatin binding, minus-end-directed, transcription co-activators. Even though the highest number of proteins corresponds to the heifer, but we found the lactation state showed the high enrichment of the ATP binding process (GO: 0005524) with p-values < 0.001 and 15.80% of involved proteins. However, RNA polymerase activities (GO: 0000978; GO: 0001228; GO: 0000977) in heifer was significant with p-values < 0.001 and proteins involvement of 4.79%. All the pregnancy-specific molecular function ontologies have non-significant p-values except the extracellular matrix structural constituent (GO: 0005201). It is also significant in lactation with p-values < 0.001. The cellular component analysis showed the identification of the different class of organelles and cellular parts in the ontological information with respective p-values (Fig. 6C). Again, we found a significant p-value < 0.001 in the lactation stage for collagen trimers (GO: 0005581). Altogether these results support our finding for the crucial importance of collagen proteins specific endogenous peptides secretion during pregnancy and lactation in which extensive remodeling of tissue and organs takes place.

Furthermore, the comparison of all three stages for the biological process identified the five most enriched ontologies (Fig. 7A). In correspondence to CC and MF process, we found the identification of extracellular matrix organization (GO: 0030198) and regulation of RNA transcription process (GO: 0000122; GO: 0045944). However, homophilic cell adhesion (GO: 0007156) and T cell differentiation (GO: 0033077) are the unique ontologies determined. Only pregnant specific ontology in homophilic cell adhesion was determined to be significant (p < 0.001). Next, we searched our list of proteins from different stages with the induction process (Fig. 7B).

We next sought to retrieve information from the urine proteome by comparing various database based on developmental stage and tissue specificity (Fig. 7C and D). We found the identification of the relevant proteins in

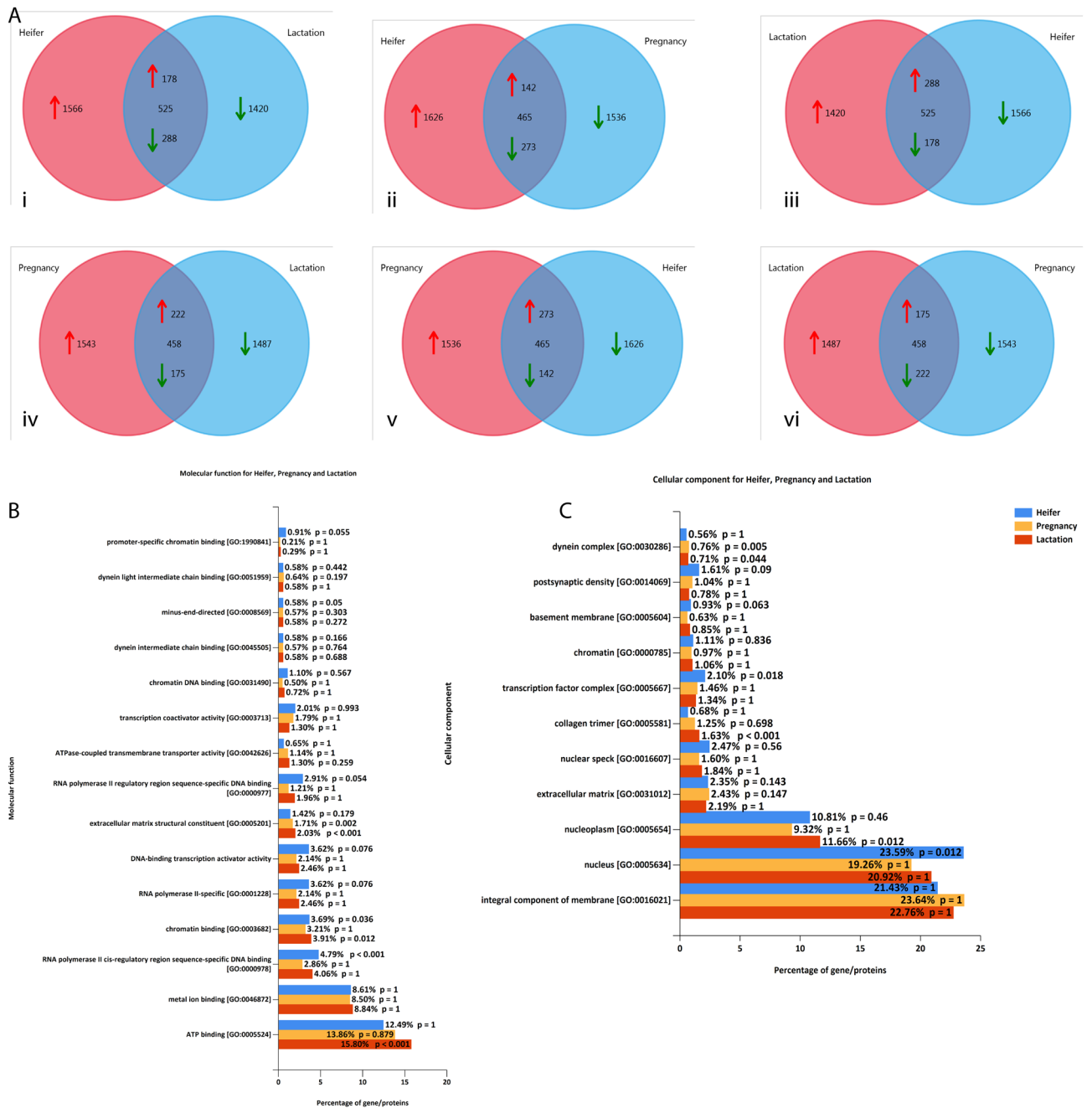


Figure 6. Parental proteome comparison: (A) Venn diagram comparison. i) Heifer versus lactation, ii) Heifer versus pregnancy, iii) Lactation versus heifer, iv) Pregnancy versus lactation, v) pregnancy versus heifer, vi) Lactation versus pregnancy. (B) Combined molecular function analysis of three conditions. (C) Cellular component analysis.

the urine specific to the stages previously reported in the literature. The results strongly suggest that the diversity of endogenous peptides is indicative of the physiological state.

Discussion

The composition of urine is an imperative indicator of the physiology of the organism and hence, can be used as an excellent non-invasively collected biofluid for diagnostic purposes. Endogenous urinary peptides have an obvious edge over full-length proteins as the formers get easy access to urine and at the same time can insinuate toward a particular protein that is disturbed within the body. The study reports a comprehensive profile of endogenous peptides in bovine urine across various physiological states and discusses their biological activities, chromosomal mapping, and molecular features. To date, this is the largest comprehensive bovine urinary peptidome dataset signifying its importance in physiology.

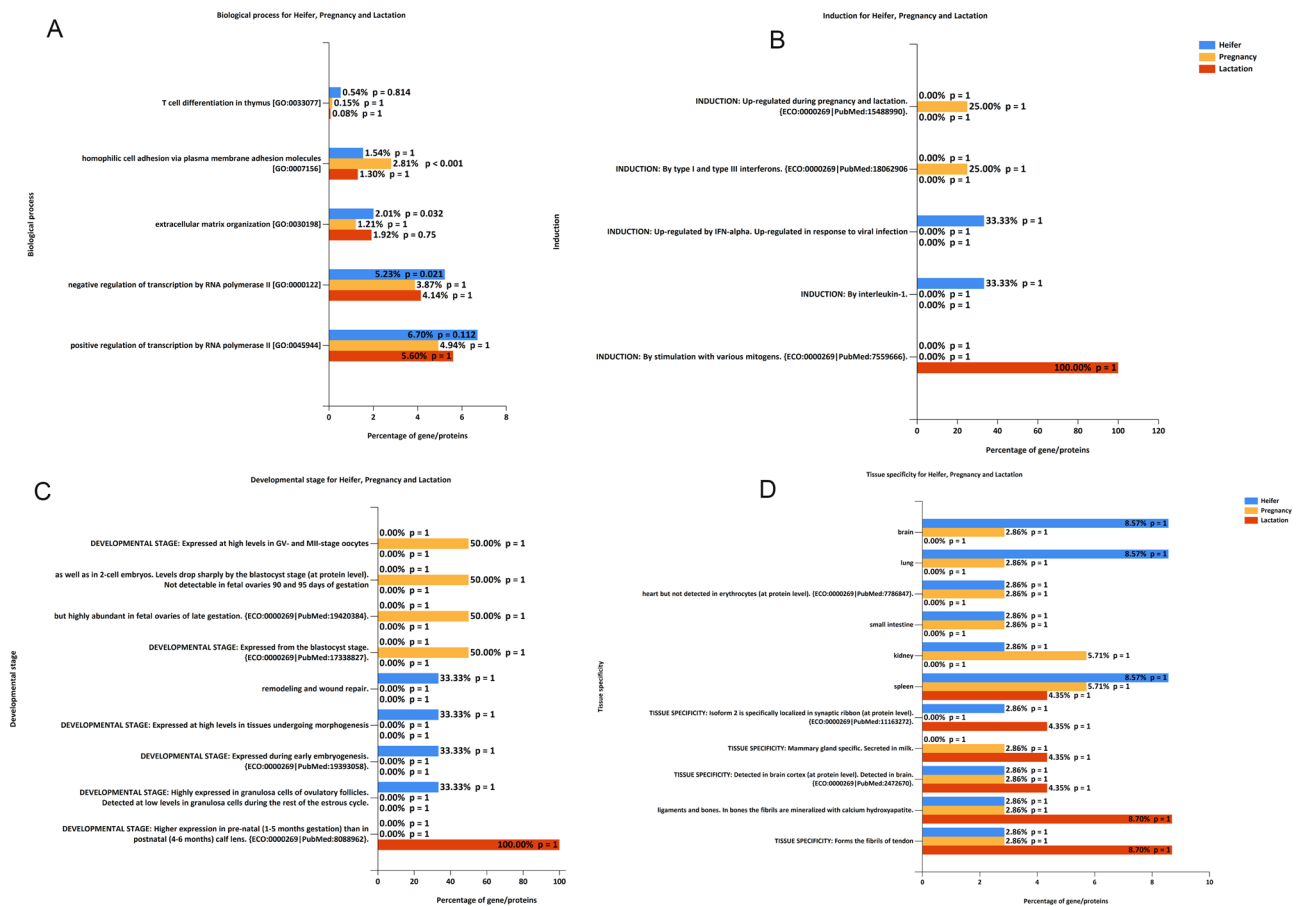


Figure 7. (A) Biological process analyses (B) Induction of proteins responsive to specific conditions. (C) Developmental proteins analysis in all three conditions. (D) Tissue-specific proteins analysis in three physiological conditions.

The urinary bladder serves as a temporary reservoir of urine. The protease activity is an essential requirement for continuous turnover of proteins in organs like urinary bladder, uterus or mammary gland wherein muscular elasticity is needed for continuous expansion and contraction. These proteases function in an interactive manner where multiple target proteins interact with multiple proteases of various families and classes creating a protease web¹⁹. Many of these proteases can be traced back to the urinary bladder which is highly stretchable. However, it is not certain whether these peptides get access to urine through cellular exudates or systemic circulation. A large number of peptides are breakdown products of large proteins involved in urine storage and voiding viz. collagen.

The molecular size distribution of the urinary peptides follows a bell-shaped curve which indicates that the majority of the peptides present in urine are between 700 to 4799 Da. Our study focuses on identification of small sized endogenous peptides (<10 kDa), by scanning MS spectra ranging from 200 to 2200 m/z, so, a fraction of very large and very small peptides might not have been covered in the study.

For peptides to get access to urine through glomerular filtrate, the permissible size cut-off is a critical determinant on the length of urinary peptides found in the urine. Of course, reabsorption in the urinary tract may affect the size distribution²⁰. In this case, it seems that most of the peptides are released into the urine because of its excess in the circulation over and above the tubular saturation limit. The observation of several peptides in urine which cannot be attributed to local organs like UB, ureter, urethra, local glands, etc is suggestive of its origin being glomerular filtrate from the blood. For example beta casein isoform and Complement C4-like isoform X1, proteins are mostly found in mammary gland and blood cells respectively²¹. The prevalence of specific peptide sequences over others suggests the crucial role executed by their precursor proteins in a specific physiological context.

To translate the results, we characterized protease substrate cleavages as net outcomes of complex proteolytic activities in normal vis-à-vis diseased physiology to foster the characterization of numerous peptide-binding receptors (PBRs) as potential drug targets. The result suggested that endogenous peptides possess a variety of bioactive functions. Prediction platforms were used to gain a comprehensive insight into the peptide sequence characteristics. We found a pattern of amino acid residues in the sequences that might be contributing to the different types of activities associated with the peptides. Many web-based prediction platforms use the Amino Acid Composition (AAC) based algorithm and literature supported the bioactivity of peptides. Searching for common residues and motifs on peptide may help us to explain the compositional biases of peptides for certain

amino acid residues and various beneficial effects attributed to the associated bioactivity. Additionally, functional assays must be performed to support the in-silico results as these algorithms might not be fully accurate.

The cow urine has been granted US Patents for its medicinal properties, particularly as a bio enhancer along with antibiotics, antifungal and anticancer drugs (6896907, 6410059). Jain *et al.* studied the effect of cow urine therapy on various types of cancers in the Mandsaur area In India wherein they reported that the severity of symptoms (pain, inflammation, burning sensation, difficulty in swallowing, and irritation) subsided significantly²². The study identifies a large number of peptides having sequence motifs unique to anticancerous properties. The Pro residue can be considered a common residue on the peptides from two categories namely anticancer and antihypertensive peptides. Scanty information is available on anticancer and antihypertensive peptides. A report determined the possible role of Ang II in tumor progression, as antihypertensive peptides target angiotensin-converting enzyme, they might be a potent anticancer agent^{23, 24}.

Many antimicrobial peptides exhibit anti-inflammatory features, the reason why we get a relatively large number of shared sequences between these two sets^{25, 26}. Wei *et al.* pointed out that many AMPs are capable of binding to LPS and this might be a possible reason for how AMPs exhibit anti-inflammatory activity by inhibiting LPS induced NO release, a pro-inflammatory mediator²⁷.

Sequence analysis showed that Gly was a common and most abundant residue in all classified peptides. Five amino acids namely Gly, Ser, Ala, Leu, and Pro in decreasing order of abundance constitute the major percentage (roughly about 50%) of all residues present in urinary peptides while in an estimate drawn from yeast proteome (6225 known and predicted proteins) it has been reported that four amino acids leucine, serine, lysine, and glutamic acid are the most abundant amino acids, totaling 32 percent of all the amino acid residues in a typical protein²⁸. Thus, it can be understood that the urinary peptides are more selective than being random. Shoom-buatong *et al.* reported that Gly is the frequently occurring residue in antibiofilm peptides (ABPs), anticancer peptides (ACPs), antifungal peptides (AFPs), antiparasitic peptides (APPs), and antiviral peptides (AVPs) at 10.98%, 10.88%, 10.79%, 10.77%, and 11.82% respectively²⁹.

ACE inhibitor peptides are reported to possess aromatic amino acids such as tryptophan, tyrosine, or proline at C-terminal^{30, 31}. Consistent with the predicted list of peptides, an adjacent proline residue to C-terminal proline residue also increases the ACE inhibitory activity of peptides³².

Anti-inflammatory peptides (AIP) exert their effects by a variety of mechanisms *viz.* inhibition of JAK-STAT and NF- κ B pathways, inducing the secretion of IL-4, inhibition of LPS induced cytokines³³⁻³⁶. Studies have reported that hydrophobic residues like phenylalanine and leucine have a major influence on the anti-inflammatory effect of peptide^{37, 38}. Manavalan *et al.* reported that average composition Arg, Leu, and Lys, were dominant in AIP³⁹. Also, Leu alone as a residue has been shown to exert anti-inflammatory activity by reducing the LPS induced NO, a pro-inflammatory mediator in RAW 264.7 macrophage. In our study, we did find Leu to be a relatively abundant amino acid followed by Gly and Ser residues.

Studies have shown that ACPs are mostly dominated by Gly, Lys, Cys, Phe, Ile, and Trp as compared to non-ACPs^{40, 41}. On contrary, in our case, sequence logo creation showed that anticancer peptide sequences are dominated by Pro, Gly, Ser, Leu, Arg residues with Pro dominating at almost every position of the sequence. A study reported that the presence of Pro residue enhances the toxicity of peptides against nucleated cancer cells while simultaneously protecting RBCs from lysis⁴². Just like AMP, Anti Cancer Peptide (ACP) also contains cationic residues. The selective killing is mediated through the electrostatic interaction between cationic residues and the anionic membrane of cancer cells⁴³. In our finding, Arg residue was the only abundant cationic residue followed by histidine. As already mentioned, collagen-derived peptides constitute a major portion of cow urinary peptides, therefore, finding Pro residue in significant abundance in the case of AHTPs and ACPs is not surprising. The wide spectrum of bioactive properties of cow urine may be associated with the small size of endogenous peptide.

Anti-Microbial Peptides are usually short (10 to 50 amino acids) with an overall positive charge (predominantly +2 to +9) and are present in every form of life⁴⁴. AMPs/ ABPs exploit the difference between the composition of the prokaryotic and eukaryotic membranes. Lipids in the outer leaflet of the animal cell contain no net charge, while prokaryotic membranes are rich in anionic phospholipids, hence, it is selectively targeted by AMPs⁴⁵.

Peptide sequences were also evaluated based on their physicochemical properties as a fine balance between charge and hydrophobicity drives the activity of the peptide. In agreement with our findings, one study finding showed that most of the position inAMPs was dominated by Arg, Lys, Leu, and Gly⁴⁶. The balance between a positive charge and hydrophobic residue drives the activity of AMP. Positively charged residues in AMP undergo electrostatic interaction with the negatively charged prokaryotic membrane, while hydrophobic residues help the AMP penetration and disruption of the bacterial membrane^{47, 48}. Chang *et al.* reported that a higher percentage of glycine and lysine was present in the core of AMP and the critical region of AMPs contained glycine as the most abundant residue⁴⁹.

The presence of collagen peptides in the urine is probably due to excessive bladder activity. Several studies reported its role in urinary bladder compliance, proliferation, and bladder filling^{50, 51}. A study shows that the ratio of type III to type I collagen determines changes in compliance in both developing fetus and mature bovine UB. Significant reduction in volumetric densities of type I and III collagen was observed with the aging of the urinary bladder in Wistar rats. The urinary bladder and other parts of the urinary system always remain in dynamic states of distension and relaxation to process and accommodate the varying volumes of urine. To ensure this, urinary epithelium and stromal cells undergo extensive extracellular matrix (ECM) degradation and remodeling. ECM epitopes, a direct resultant of MMP mediated proteolysis, are involved in the stretch-induced proliferation of bladder smooth muscle cells through ERK1/2 signaling activation⁵². Collagen being a connective tissue is the major target in the ECM degradation and remodeling process and hence shows its signature in urine in the form of various small-sized EPs. Reinforced by other reports, our finding shows that collagen-derived peptides are abundant in urine^{10, 53, 54} but the individual sequences of the peptides were different in different physiological

states. One study found a significant correlation of the relative abundance between urine and plasma samples and hypothesized that collagen-derived peptides, unlike any other peptide in plasma, by an unknown mechanism, escapes tubular reabsorption⁵⁴. In urine, the peptide degradation is least in comparison to other biological fluid like blood serum or plasma⁵⁵. Thus, making it a suitable biological fluid for proteome-based study.

The role of proteases is important for the complete turnover of proteins in various physiological and diseases condition. Approximately, 5–10% of drugs target these proteases⁵⁶. Most of the peptides in the biological fluid are a consequence of proteolytic cleavage of proteins by diverse types of proteases. The disturbance in the protease network leads to various diseases including cancer, cardiovascular, inflammatory, and gastrointestinal disorders^{57–59}. However, so far technical limitations have prohibited a global understanding of interconnected protease activity in complex pathophysiological situations. Since pre-pregnancy (heifer), pregnancy and lactation are distinct physiological states; we traced the urinary peptide profile back to the systemic changes in the protease network with a view to identifying any significant correlation between specific protease and physiology. A significant finding was the kallikrein and matrix metalloproteinases class of proteases were actively involved in proteolytic events. MMP class of proteases serves diverse roles; most of them are associated with collagen degradation, cell migration, cancer progression, collagen affinity enhancing, pro and anti-inflammatory activities, etc. Proteases like plasmin, trypsin, chymase, and certain other metalloproteinases viz. MMP-1, MMP-2, MMP-8, MMP-9 are involved in the activation of MMPs⁶⁰.

We identified peptides of Matrix Metallo Proteases (MMP proteases) in urinary peptides (Heifer: MMP-2, 9, 11, 13, 16, 0, 23, 25, 27, 28; Lactation: MMP-9, 16, 17, 23, 24, 25; Pregnant: MMP-8 & 20), which supports the proposed role of MMPs in ECM remodeling of genito-urinary tract and explains the origin of collagen-derived peptides in urine.

Besides MMPs, cathepsin proteases were also predicted with quite a significant number of cleavages and we also found their signatures in MS/MS data. Cathepsins are lysosomal enzymes which include serine, cysteine, and aspartic proteases, just like MMPs, these are also involved in the remodeling of ECM^{61,62}.

Interestingly, during pregnancy, the resultant peptides from protease cleavage were significantly less as compared to the other physiological states. Similar observation was made in the study in mice where the activity of certain proteases such as MMP2 and MMP9 is suppressed during pregnancy, which increases after parturition and reverts to normal⁶³.

Molecular Function ontology: extracellular matrix structural constituent (GO:0005201) was found significant in the lactation and pregnancy indicating tissue remodeling during these stages. COL8A1 gene was found upregulated during tissue remodeling in pregnant myometrium (human)⁶⁴. During mammary cell growth in bovine, COL8A1 is involved in epithelial cell proliferation which was found to be upregulated during lactation⁶⁵. Homophilic cell adhesion (GO: 0007156) was found to be significant in pregnancy stage which plays significant role during embryo implantation^{66,67}.

Searching list of proteins from different stages with induction process showed that none of the processes was found to be shared in all three stages; however, proteins associated with specific processes were explicitly relevant to the physiological condition. We identified the GLUT8 peptides in pregnant animal urine, which was upregulated during the pregnancy induction process. GLUT8 mRNA expression in the bovine mammary gland was increased 10-fold ($P < 0.001$) during late pregnancy and early lactation⁶⁸.

Conclusions

We presented a simple method for the discovery of thousands of endogenous peptides in cow urine that contribute to various bioactivities associated with urine. The data presented here represents the identification of 5000 natural peptides in all three physiological conditions laying the foundations for detailed studies. The molecular weight, peptide length, and amino acid distribution of endogenous peptides follow a similar pattern in all three stages. The study emphasized on the identification and characterization of endogenous peptides of <10 kDa molecular weight by scanning MS spectra ranging from 200 to 2200 m/z. which represented peptides of molecular weight ranging from 700 to 4799 Da. As our study used C18 silica gel for the peptide extraction, many of the highly cationic and hydrophilic peptides might not have been covered. A combination of chromatographic methods can still improve the identification coverage of urinary peptides. Several of the bioactive properties associated with the peptides were predicted using in-silico platforms. We provided evidence for the peptide-mediated antimicrobial activity against *E. coli* and *S. aureus* but more experiments are needed to validate other predicted bioactivities. This study also analyzed the complex network of proteases active during specific physiological states and the target proteins as precursors of the urinary peptides. The knowledge about physiology-specific proteins and proteases may generate further interest in the field of biomarker discovery for pregnancy diagnosis or in understanding the pathophysiology during development and lactation. The urinary peptides also represent degradome and explain the formation and release of peptides from precursor proteins by specific proteases. Lastly, the tissue-specific developmental gene ontological classification derived from the protease-targeted proteome highlighted the stages at which specific set of proteins are active in specific physiology. The data opens future avenues to set a benchmark for urine-based biomarkers unaffected by physiology and also helps to understand various functional activities associated with cow urine. The present study has searched for the peptides based on primary sequence molecular weight only. However, it will be interesting to observe and search for post-translationally modified peptides including glycosylated, phosphorylated, and others.

Methods

Sample collection. The urine samples were collected from 30 healthy female Sahiwal cattle [*Bos indicus*, belonging to three different physiological states namely, heifer ($n = 10$), pregnancy ($n = 10$), and lactation ($n = 10$) in the presence of a veterinary doctor from Livestock Research Centre located at National Dairy Research

Institute (NDRI) Karnal. All the animals included in the study were clinically healthy and divided into three groups (Heifer (age between 17th–18th months), pregnant (40th–60th days of pregnancy), and lactating (80th–100th days of lactation)). All procedures were approved by the Institutional Animal Ethics Committee (IAEC) ICAR-NDRI, Karnal, India.

Within each group, a sample was created by pooling samples of ten animals (biological replicates $n = 10$) and processed separately with six technical replicates. However, to understand the animal to animal biological variations, we performed a separate study on seven individual animals from each category. The procedure for the samples collection and processing is the same for all the groups as described here briefly. Approx. 500 mL of fixed time morning voids urine samples were collected by massaging the perineum of the animal manually. The collected samples were immediately transferred to the lab and analyzed for any debris, dung, dust, or hair to rule out any contaminants in the samples. Initially, the samples were filtered by muslin cloth followed by centrifuging at 7000 rpm for 20 min to allow settlement of any cell debris and particulate matter. The microscopic examination was performed for individual samples, before and after the centrifugation, to observe the presence of RBCs, WBCs, other cells, and debris. The purified clean urine was further used for peptide extraction and purification (Fig. 1A).

Peptide extraction. The clean urine samples were passed through ultra-filtration assembly (Thermo easy-load Masterflex, model 7518-00, USA) with a 10 kDa molecular weight cut-off filter (Pall Minimate™ TFF Capsule), to separate the endogenous peptides mixture. The pH of the obtained filtrate was adjusted to ≤ 3 using Trifluoroacetic acid (TFA) for further treatment. The peptides mixture (PM) was subjected to manually prepared C-18 beads based Solid Phase Extraction columns (SPE) (Fig. 1B). Briefly, the column was prepared using the C-18 reversed-phase silica gel (Sigma, Switzerland Cat. No.60757-50G). A slurry of 20 grams of silica gel was prepared in methanol. The packing of the column was done by continuously stirring the slurry and then slowly draining it into the column followed by several washes with methanol.

The packed column was conditioned using 90% methanol followed by equilibration with 10 column volumes of 0.1% TFA in water. After equilibration, PM was loaded with a flow rate of 0.5 ml/min followed by desalting using 5% methanol with 0.1% TFA. The desalted peptides were eluted in 60% acetonitrile (ACN) with 0.1% TFA. Processing of around 500 mL of urine yielded approx. 30–50 mL of eluate with dark brown appearance. For the removal of the dark brown substances (possible contaminating metabolites in urine) the ethyl acetate-based extraction was performed. The eluates were subjected to the double volume of ethyl acetate followed with end to end rotation for 5 min and allowed to settle for 10 min to differentiate into two layers. The upper organic layer was stored appropriately for metabolome profiling and the lower aqueous phase was aliquoted in 2 mL microcentrifuge tubes and dried by speed vac (Thermo savant ISS110 SpeedVac concentrator, ISS110-230, USA).

The dried samples were stored at -80°C . Prior to mass spectrometer run the routine sample cleanup was performed by Pierce C-18 Spin Columns as per the manufacturer's protocol. Briefly, the quantified samples were reconstituted in 20% ACN with 2% TFA (sample buffer, 3 μL for every 1 μL of the sample). After loading and washing of the sample, peptides were eluted in 70% ACN with 0.1% TFA. The eluted samples were dried and reconstituted in 0.1% TFA before MS run.

Electrospray ionization tandem mass spectrometry LC-MS/MS analysis. The reconstituted peptides were used for shotgun proteomics experiments for the identification of the endogenous peptides. The peptides were separated using micro-LC (Dionex, Thermo UltiMate 3000 HPLC System, USA) through analytical column (Supelco, Ascentis Express C18, 25 cm \times 4.6 mm, 2.7 μm) coupled with ESI source (BrukerDaltonics, Germany) spray in Maxis-HD qTOF (Bruker, Germany) mass spectrometer. The acquisition parameters were adapted from our previous reports with slight modifications^{69,70}. The elution was performed with a flow rate of 150 $\mu\text{L}/\text{min}$ a continuous gradient of 5–75% acetonitrile over 135 min. In the solvent system; Solvent A was 100% water with 0.1% formic acid, and solvent B was 100% acetonitrile with 0.1% formic acid. Data were acquired in the data-dependent mode in mass spectrometer operated in automatically switching between MS and MS/MS acquisition. The precursor ion MS spectra scan range of 200–2200 (m/z) was used in the Q-TOF with resolution $R = 75,000$. The six most abundant precursor ions were searched for detection of different masses during acquisition and selected for fragmentation using collision-induced dissociation (CID) with a fixed cycle time of 3 s along with 2 min of release for exclusion filter (otof processing software, BrukerDaltonics).

Data processing. The .d raw data files were subjected to TPP pipeline for the identification of endogenous peptides. For analysis, the tof generated raw (.d) files were converted to mzML format using MSconvert GUI using the default parameters. The converted mzML files were searched for MS/MS spectra using the Trans-Proteomic Pipeline version 5.1.0 released on 2017-11-03 on in-house combined UniProt *Bos taurus* (Cow), *Bubalus bubalis* (Buffalo), common contaminant sequences, and an equal number of decoy sequences database. The detailed protocol was reported in Suhail et al., 2019⁷¹. Briefly, for the analysis, the peptide assignments were performed using multiple search engines using X! Tandem (with the k-score plug-in), and Comet. For both the search engine the search parameters included un-digested peptides and the remaining parameters were kept as default. The minimum peptide length parameter was set to 6 amino acid residues. Further Peptide Prophet and Protein Prophet algorithms were used in the pipeline to compute the probabilities score for both individually searched peptides and the respective proteins. The accurate mass model in Peptide Prophet was used for high confidence peptide identifications to boost the probability of peptide identification. Another protein validation step was executed using both Peptide Prophet and Protein Prophet Scores, where the protein was authenticated if it contained a minimum of two top-ranked peptides with each peptide probability score above 95%. All the search engine results were merged and validated using iProphet. This method takes the input of Peptide

Prophet spectrum-level results from multiple LC-MS/MS runs and then computes a new probability at the level of a unique peptide sequence (or protein sequence). This framework allows for the combination of results from multiple search tools and takes into account other supporting factors, including the number of sibling experiments identifying the same peptide ions, the number of replicate ion identifications, sibling ions, and sibling modification states. A model of iProphet performance concerning the number of correct identifications versus error. An iProphet probability of more than 0.95 was used as the cutoff for the final identification of the protein. For protein quantitation, the MS2-based Normalized Spectral Index (SIN) was used in the StPeter algorithm implemented in TPP for all proteins identified with ≥ 2 unique peptides per protein.

Antimicrobial assay. The total urinary peptide from all three groups extracted by SPE was then assessed for its a AntiMicrobial Activity (AMA). To determine its AMA the peptide was reconstituted in milliQ water and coated on 6 mm sterile discs (HIMEDIA, SD067-1VL). The discs were allowed to dry under a laminar hood. A 0.5 McFarland equivalent of test cultures (*Staphylococcus aureus* ATCC 29213, *Escherichia coli* ATCC 25922) were swabbed on the surface of Mueller Hinton Agar (HIMEDIA, GM173-500G) and allowed to dry. With the help of sterile forceps, the coated discs were placed on the lawn of the test culture and incubated overnight. BSA digest was used as the negative control in the experiment. The appearance of the inhibition zone confirms the antimicrobial activity of the cow urinary peptides derived from all physiological states.

Bioinformatics analysis. All the graphical analyses were performed in the R environment using the ggPlot2 package⁷². The Gene Ontology (GO) categories were analyzed using the DAVID bioinformatics resources and only the genes with an adjusted P-value (false-discovery rate) of less than 0.05 were included for subsequent GO term plotting.

Strategy for manual and MERPOS based protease prediction. Peptide list belonging to different physiological states of cows was sorted based on possible protease cleavage site located on four residues either from N or C terminal. The hypothesis was to identify the protease responsible for its release from the precursor sequence. The peptide sequences were traced back to 21 possible proteases, which provided hints to explain the physiology driven variation in the number and type of sequences derived from each physiological state.

Proteasix tool was used for the prediction of cleavage sites which uses the MEROPS peptidase database as a reference. It provides specificity weight matrices for experimentally validated cleavages for proteases. The batch peptide match tool from Protein Information Resource (PIR) was used to retrieve the start and end amino acid position from the peptide sequence. The retrieved information from PIR was used as an input for Proteasix. Cleavages predicted with more than 80% specificity were selected for the heat map analysis. The aforementioned method was used to predict the cleavages and the associated proteases for heifer, lactation, and pregnant groups.

Characterization and classification of bioactive peptides based on machine learning SVM model.

In-silico analysis was done using web-based platforms to identify bioactive peptides present in cow urine. For validation purposes, the online activities predicting servers using Support Vector Machine (SVM) based algorithm were used. SVM, an established machine learning tool has been extensively used to design prediction platforms for the classification of various kinds of bioactivities associated with peptides^{73–76}. Peptides sequences associated with anticancer, anti-inflammatory (AIP), antihypertensive, antimicrobial, and antibiofilm activities were predicted with a high confidence score at or above a higher threshold value of 0.9. Peptides with ACE-inhibitory effect of antihypertensive activity were obtained from antihypertensive peptides inhibitors (AHTpin) (<http://crdd.osdd.net/raghava/ahtpin/>). The database consists of 1745 entries as a positive dataset from various sources viz. AHTPDB, BIOPEP, ACEpepDB, and literature whereas negative datasets consist of a random fragment of proteins of the same length from Swiss-Prot⁷⁵. Sequence logo was designed for the predicted sequence retrieved from AHTpin to determine the abundance of amino acid residues at the C-terminus of the peptides as the presence of certain amino acid residues has a significant impact on antihypertensive activity. For the prediction of AIPs, a web server anti-inflam (<http://metagenomics.iiserb.ac.in/antiinflam>) was accessed.

Anticancer activity was predicted using the tumorHPD server (<http://crdd.osdd.net/raghava/tumorhpd/>). This tool utilizes 651 experimentally validated peptides (peptides binding to tumor) in a positive dataset and 651 non-tumor binding peptides randomly generated from proteins obtained from SwissProt74.

Two servers CAMPR3 (<http://www.camp3.bicnirrh.res.in/campHelp.php>), a database containing entries of antimicrobial peptides from various sources and the other dPABBs which uses SVM based algorithm for the prediction of Anti Biofilm Peptides (ABPs) (<http://ab-openlab.csir.res.in/abp/antibiofilm/>) were used to predict antimicrobial activity associated with the urinary peptides. Peptides exhibiting antibiofilm activity on at least 4 prediction models were selected. The tool uses a positive dataset consisting of 90 AMPs that are proven through in vitro and in vivo reports to be active against biofilms and 88 Quorum sensing peptides constitute negative dataset. The prediction is based upon whole amino acid composition, features of the residues, and their positional preferences⁷⁶.

Chromosomal mapping of urinary peptides. All the identified proteins were mapped with Uniprot bovine chromosomal proteome information database (*Bos taurus* taxid 9931) and parsed to create files appropriately formatted for input to Circos for circular visualization and annotation⁷⁷.

Ethical approval and consent to participate. The approval for the experiment was obtained from Institutional Animal Ethics Committee (IAEC) of National Dairy Research Institute (IAEC approval no. 41-IAEC-

18–50). Methods were carried out in accordance to IAEC and ARRIVE guidelines. The urine samples were collected from the Livestock Research Centre (LRC) of National Dairy Research Institute (NDRI), Karnal which is a public funded research institute under the Indian Council of Agricultural Research, Government of India. The second morning void urine samples were collected from three different physiological states of Sahiwal cow. The animals were maintained under expert veterinary supervision. NDRI has all the necessary permits for the housing and care of animals for scientific purposes vide registration no. 1705/GO/ac/13/CPCSEA 3rd July, 2013 duly approved by Ministry of Environment and Forest, Govt. of India (Web site: <http://envfor.nic.in>).

Received: 18 December 2020; Accepted: 4 May 2021

Published online: 14 June 2021

References

1. Tebbe, A. W. & Weiss, W. P. Evaluation of creatinine as a urine marker and factors affecting urinary excretion of magnesium by dairy cows. *J. Dairy Sci.* <https://doi.org/10.3168/jds.2017-14098> (2018).
2. Schaub, S. *et al.* Urine protein profiling with surface-enhanced laser-desorption/ionization time-of-flight mass spectrometry. *Kidney Int.* **65**, 323–332 (2004).
3. Bauça, J. M., Martínez-Morillo, E. & Diamandis, E. P. Peptidomics of urine and other biofluids for cancer diagnostics. *Clin. Chem.* <https://doi.org/10.1373/clinchem.2013.211714> (2014).
4. Pisitkun, T., Johnstone, R. & Knepper, M. A. Discovery of urinary biomarkers. *Mol. Cell. Proteom.* <https://doi.org/10.1074/mcp.R600004-MCP200> (2006).
5. Kalantari, S., Jafari, A., Moradpoor, R., Ghasemi, E. & Khalkhal, E. Human urine proteomics: Analytical techniques and clinical applications in renal diseases. *Int. J. Proteom.* <https://doi.org/10.1155/2015/782798> (2015).
6. Belczacka, I. *et al.* Urinary CE-MS peptide marker pattern for detection of solid tumors. *Sci. Rep.* <https://doi.org/10.1038/s41598-018-23585-y> (2018).
7. Hirao, Y. *et al.* Proteome profiling of diabetic mellitus patient urine for discovery of biomarkers by comprehensive MS-based proteomics. *Proteomes* <https://doi.org/10.3390/proteomes6010009> (2018).
8. Arthur, J. M., Karakala, N. & Edmondson, R. D. Proteomic analysis for identification of biomarkers that predict severe acute kidney injury. *Nephron* <https://doi.org/10.1159/000491440> (2018).
9. Mariani, L. H. *et al.* CureGN study rationale, design, and methods: Establishing a large prospective observational study of glomerular disease. *Am. J. Kidney Dis.* <https://doi.org/10.1053/j.ajkd.2018.07.020> (2019).
10. Good, D. M. *et al.* Naturally occurring human urinary peptides for use in diagnosis of chronic kidney disease. *Mol. Cell. Proteomics* <https://doi.org/10.1074/mcp.M110.001917> (2010).
11. Smith, C. R. *et al.* Deciphering the peptidome of urine from ovarian cancer patients and healthy controls. *Clin. Proteomics* <https://doi.org/10.1186/1559-0275-11-23> (2014).
12. Stalmach, A. *et al.* Identification of urinary peptide biomarkers associated with rheumatoid arthritis. *PLoS ONE* <https://doi.org/10.1371/journal.pone.0104625> (2014).
13. Bathla, S. *et al.* Profiling of urinary proteins in Karan Fries cows reveals more than 1550 proteins. *J. Proteomics* <https://doi.org/10.1016/j.jprot.2015.05.026> (2015).
14. Rawat, P. *et al.* Identification of potential protein biomarkers for early detection of pregnancy in cow urine using 2D DIGE and label free quantitation. *Clin. Proteomics* <https://doi.org/10.1186/s12014-016-9116-y> (2016).
15. Shashikumar, N. G. *et al.* Global proteomic analysis of water buffalo (*Bubalus bubalis*) saliva at different stages of estrous cycle using high throughput mass spectrometry. *Theriogenology* <https://doi.org/10.1016/j.theriogenology.2017.12.046> (2018).
16. Pyo, J. *et al.* Characterization of a bovine pregnancy-associated protein using two-dimensional gel electrophoresis. N-terminal sequencing and mass spectrometry. *Proteomics* <https://doi.org/10.1002/pmic.200300618> (2003).
17. Bislev, S. L. *et al.* A Bovine PeptideAtlas of milk and mammary gland proteomes. *Proteomics* <https://doi.org/10.1002/pmic.201200057> (2012).
18. Chopra, A. *et al.* High-resolution mass spectrometer-based ultra-deep profile of milk whey proteome in Indian Zebu (Sahiwal) Cattle. *Front. Nutr.* **7**, 150 (2020).
19. Overall, C. M. & Kleifeld, O. Validating matrix metalloproteinases as drug targets and anti-targets for cancer therapy. *Nat. Rev. Cancer* <https://doi.org/10.1038/nrc1821> (2006).
20. Lawrence, M. G. *et al.* Permeation of macromolecules into the renal glomerular basement membrane and capture by the tubules. *Proc. Natl. Acad. Sci. U. S. A.* <https://doi.org/10.1073/pnas.1616457114> (2017).
21. Hu, S., Loo, J. A. & Wong, D. T. Human body fluid proteome analysis. *Proteomics* <https://doi.org/10.1002/pmic.200600284> (2006).
22. Dhama, K., Chauhan, R. & Singhal, L. Anti-cancer activity of cow urine: Current status and future directions. *Int. J. Cow Sci.* (2005).
23. Lever, A. F. *et al.* Do inhibitors of angiotensin-I-converting enzyme protect against risk of cancer?. *Lancet* [https://doi.org/10.1016/S0140-6736\(98\)03228-0](https://doi.org/10.1016/S0140-6736(98)03228-0) (1998).
24. Ebrahimi, L., Ai, J., Alizadeh, A. & Shariati, M. Imminent angiotensin-converting enzyme inhibitor from microbial source for cancer therapy. *Int. J. Prev. Med.* https://doi.org/10.4103/ijpvm.IJPVM_324_16 (2017).
25. Mahlapuu, M., Håkansson, J., Ringstad, L. & Björn, C. Antimicrobial peptides: An emerging category of therapeutic agents. *Front. Cell. Infect. Microbiol.* <https://doi.org/10.3389/fcimb.2016.00194> (2016).
26. Shim, D. W. *et al.* Anti-inflammatory action of an antimicrobial model peptide that suppresses the TRIF-Dependent signaling pathway via inhibition of toll-like receptor 4 endocytosis in lipopolysaccharide-stimulated macrophages. *PLoS ONE* <https://doi.org/10.1371/journal.pone.0126871> (2015).
27. Wei, L. *et al.* Structure and function of a potent lipopolysaccharide-binding antimicrobial and anti-inflammatory peptide. *J. Med. Chem.* <https://doi.org/10.1021/jm4004158> (2013).
28. Harvey Lodish, Arnold Berk, S Lawrence Zipursky, Paul Matsudaira, David Baltimore, and J. D. Section 3.1, Hierarchical Structure of Proteins. in *Molecular Cell Biology*. (W.H. Freeman & Co Ltd, 2000).
29. Shooabuatong, W., Schaduangrat, N. & Nantasenam, C. Unraveling the bioactivity of anticancer peptides as deduced from machine learning. *EXCLI J.* <https://doi.org/10.17179/excli2018-1447> (2018).
30. Maruyama, S. *et al.* Angiotensin I-Converting Enzyme Inhibitory Activity of the C-Terminal Hexapeptide of alpha s1-Casein. *Agric. Biol. Chem.* <https://doi.org/10.1080/00021369.1987.10868415> (1987).
31. Kim, Y. K. & Chung, B. H. A novel angiotensin-I-converting enzyme inhibitory peptide from human alpha(s1)-casein. *Biotechnol. Lett.* <https://doi.org/10.1023/A:1005572504196> (1999).
32. Muirhead, E. E., Brooks, B. & Arora, K. K. Prevention of malignant hypertension by the synthetic peptide SQ 20,881. *Lab. Investig.* (1974).

33. La Manna, S. *et al.* Mimetics of suppressor of cytokine signaling 3: Novel potential therapeutics in triple breast cancer. *Int. J. Cancer* <https://doi.org/10.1002/ijc.31594> (2018).
34. Kurinami, H. *et al.* A novel therapeutic peptide as a partial agonist of RANKL in ischemic stroke. *Sci. Rep.* <https://doi.org/10.1038/srep38062> (2016).
35. Zellinger, C., Salvamoser, J. D., Seeger, N., Russmann, V. & Pötschka, H. Impact of the neural cell adhesion molecule-derived peptide FGL on seizure progression and cellular alterations in the mouse kindling model. *ACS Chem. Neurosci.* <https://doi.org/10.1021/cn400153g> (2014).
36. Caceres, C. C. *et al.* An engineered cyclic peptide alleviates symptoms of inflammation in a murine model of inflammatory bowel disease. *J. Biol. Chem.* <https://doi.org/10.1074/jbc.M117.779215> (2017).
37. Nan, Y. H. *et al.* Antimicrobial and anti-inflammatory activities of a Leu/Lys-rich antimicrobial peptide with Phe-peptoid residues. *Protein Pept. Lett.* **14**, 1003–1007 (2007).
38. Vogel, H. J. *et al.* Towards a structure-function analysis of bovine lactoferricin and related tryptophan- and arginine-containing peptides. *Biochem. Cell Biol.* <https://doi.org/10.1139/o01-213> (2002).
39. Manavalan, B., Shin, T. H., Kim, M. O. & Lee, G. AIPred: Sequence-based prediction of anti-inflammatory peptides using random forest. *Front. Pharmacol.* <https://doi.org/10.3389/fphar.2018.00276> (2018).
40. Tyagi, A. *et al.* In silico models for designing and discovering novel anticancer peptides. *Sci. Rep.* <https://doi.org/10.1038/srep02984> (2013).
41. Chen, W., Ding, H., Feng, P., Lin, H. & Chou, K. C. iACP: A sequence-based tool for identifying anticancer peptides. *Oncotarget* (2016) <https://doi.org/10.18632/oncotarget.7815>.
42. Vermeer, L. S. *et al.* Conformational flexibility determines selectivity and antibacterial, antiplasmodial, and anticancer potency of cationic- α helical peptides. *J. Biol. Chem.* <https://doi.org/10.1074/jbc.M112.359067> (2012).
43. Mader, J. S. & Hoskin, D. W. Cationic antimicrobial peptides as novel cytotoxic agents for cancer treatment. *Expert Opin. Investig. Drugs* <https://doi.org/10.1517/13543784.15.8.933> (2006).
44. Wang, J. *et al.* Biodegradable hydrophilic polyurethane PEGU25 loading antimicrobial peptide Bmap-28: A sustained-release membrane able to inhibit bacterial biofilm formation in vitro. *Sci. Rep.* <https://doi.org/10.1038/srep08634> (2015).
45. Lohner, K., Latal, A., Lehrer, R. I. & Ganz, T. Differential scanning microcalorimetry indicates that human defensin, HNP-2, interacts specifically with biomembrane mimetic systems. *Biochemistry* <https://doi.org/10.1021/bi961300p> (1997).
46. Lata, S., Sharma, B. K. & Raghava, G. P. S. Analysis and prediction of antibacterial peptides. *BMC Bioinformatics* <https://doi.org/10.1186/1471-2105-8-263> (2007).
47. Dathe, M. & Wieprecht, T. Structural features of helical antimicrobial peptides: Their potential to modulate activity on model membranes and biological cells. *Biochim. Biophys. Acta Biomembranes* [https://doi.org/10.1016/S0005-2736\(99\)00201-1](https://doi.org/10.1016/S0005-2736(99)00201-1) (1999).
48. Michael Henderson, J. & Lee, K. Y. C. Promising antimicrobial agents designed from natural peptide templates. *Curr. Opin. Solid State Mater. Sci.* <https://doi.org/10.1016/j.cossms.2013.08.003> (2013).
49. Chang, K. Y., Lin, T. P., Shih, L. Y. & Wang, C. K. Analysis and prediction of the critical regions of antimicrobial peptides based on conditional random fields. *PLoS ONE* <https://doi.org/10.1371/journal.pone.0119490> (2015).
50. Chang, S. L., Howard, P. S., Koo, H. P. & Macarak, E. J. Role of type III collagen in bladder filling. *NeuroUrol. Urodyn.* [https://doi.org/10.1002/\(SICI\)1520-6777\(1998\)17:2<135::AID-NAU7>3.0.CO;2-E](https://doi.org/10.1002/(SICI)1520-6777(1998)17:2<135::AID-NAU7>3.0.CO;2-E) (1998).
51. Macarak, E. J. & Howard, P. S. The role of collagen in bladder filling. *Adv. Exp. Med. Biol.* https://doi.org/10.1007/978-1-4615-4737-2_17 (1999).
52. Aitken, K. J. *et al.* Mechanotransduction of extracellular signal-regulated kinases 1 and 2 mitogen-activated protein kinase activity in smooth muscle is dependent on the extracellular matrix and regulated by matrix metalloproteinases. *Am. J. Pathol.* <https://doi.org/10.2353/ajpath.2006.050969> (2006).
53. Coon, J. J. *et al.* CE-MS analysis of the human urinary proteome for biomarker discovery and disease diagnostics. *Proteomics Clin. Appl.* <https://doi.org/10.1002/prca.200800024> (2008).
54. Magalhães, P. *et al.* Comparison of urine and plasma peptidome indicates selectivity in renal peptide handling. *Proteomics Clin. Appl.* <https://doi.org/10.1002/prca.201700163> (2018).
55. Good, D. M. *et al.* Body fluid proteomics for biomarker discovery: Lessons from the past hold the key to success in the future. *J. Proteome Res.* <https://doi.org/10.1021/pr070529w> (2007).
56. Drag, M. & Salvesen, G. S. Emerging principles in protease-based drug discovery. *Nat. Rev. Drug Discov.* <https://doi.org/10.1038/nrd3053> (2010).
57. Mason, S. D. & Joyce, J. A. Proteolytic networks in cancer. *Trends Cell Biol.* <https://doi.org/10.1016/j.tcb.2010.12.002> (2011).
58. Sharony, R. *et al.* Protein targets of inflammatory serine proteases and cardiovascular disease. *J. Inflamm.* <https://doi.org/10.1186/1476-9255-7-45> (2010).
59. Antalis, T. M., Shea-Donohue, T., Vogel, S. N., Sears, C. & Fasano, A. Mechanisms of disease: Protease functions in intestinal mucosal pathobiology. *Nat. Clin. Pract. Gastroenterol. Hepatol.* <https://doi.org/10.1038/ncpgasthep0846> (2007).
60. Jabłońska-Trypuć, A., Matejczyk, M. & Rosochacki, S. Matrix metalloproteinases (MMPs), the main extracellular matrix (ECM) enzymes in collagen degradation, as a target for anticancer drugs. *J. Enzyme Inhibit. Med. Chem.* <https://doi.org/10.3109/1475366.2016.1161620> (2016).
61. Sobotič, B. *et al.* Proteomic identification of cysteine cathepsin substrates shed from the surface of cancer cells. *Mol. Cell. Proteomics* <https://doi.org/10.1074/mcp.M114.044628> (2015).
62. Gocheva, V. & Joyce, J. A. Cysteine cathepsins and the cutting edge of cancer invasion. *Cell Cycle* <https://doi.org/10.4161/cc.6.1.3669> (2007).
63. Wieslander, C. K. *et al.* Regulation of elastolytic proteases in the mouse vagina during pregnancy, parturition, and puerperium. *Biol. Reprod.* <https://doi.org/10.1095/biolreprod.107.063024> (2008).
64. Rehman, K. S., Yin, S., Mayhew, B. A., Word, R. A. & Rainey, W. E. Human myometrial adaptation to pregnancy: cDNA microarray gene expression profiling of myometrium from non-pregnant and pregnant women. *Mol. Hum. Reprod.* <https://doi.org/10.1093/molehr/gag078> (2003).
65. Dai, W., Zou, Y., White, R. R., Liu, J. & Liu, H. Transcriptomic profiles of the bovine mammary gland during lactation and the dry period. *Funct. Integr. Genomics* <https://doi.org/10.1007/s10142-017-0580-x> (2018).
66. Fukuda, M. N. & Nozawa, S. Trophinin, tasin, and bystin: A complex mediating unique attachment between trophoblastic and endometrial epithelial cells at their respective apical cell membranes. *Sem. Reprod. Endocrinol.* <https://doi.org/10.1055/s-2007-1016230> (1999).
67. Sugihara, K. *et al.* Induction of trophinin in human endometrial surface epithelia by CGbeta and IL-1beta. *FEBS Lett.* <https://doi.org/10.1016/j.febslet.2007.12.001> (2008).
68. Zhao, F. Q. *et al.* Bovine glucose transporter GLUT8: Cloning, expression, and developmental regulation in mammary gland. *Biochim. Biophys. Acta Gene Struct. Expr.* <https://doi.org/10.1016/j.bbaexp.2004.09.001> (2004).
69. Ali, S. A., Singh, P., Tomar, S. K., Mohanty, A. K. & Behare, P. Proteomics fingerprints of systemic mechanisms of adaptation to bile in *Lactobacillus fermentum*. *J. Proteomics* **213**, 103600 (2020).
70. Ali, S. A. *et al.* Examination of pathways involved in leukemia inhibitory factor (LIF)-induced cell growth arrest using label-free proteomics approach. *J. Proteomics* **168**, 37–52 (2017).

71. Suhail, A. *et al.* DeSUMOylase SENP7-mediated epithelial signaling triggers intestinal inflammation via expansion of gamma-delta T cells. *Cell Rep.* <https://doi.org/10.1016/j.celrep.2019.11.028> (2019).
72. Wickham, H. *Elegant Graphics for Data Analysis* (Springer, Berlin, 2016).
73. 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. Bioinforma.* <https://doi.org/10.1109/TCBB.2012.89> (2012).
74. Sharma, A. *et al.* Computational approach for designing tumor homing peptides. *Sci. Rep.* <https://doi.org/10.1038/srep01607> (2013).
75. Kumar, R. *et al.* AHTPDB: a comprehensive platform for analysis and presentation of antihypertensive peptides. *Nucleic Acids Res.* **43**, 956–962 (2015).
76. Sharma, A., Gupta, P., Kumar, R. & Bhardwaj, A. DPABBS: A novel in silico approach for predicting and designing anti-biofilm peptides. *Sci. Rep.* <https://doi.org/10.1038/srep21839> (2016).
77. Krzywinski, M. *et al.* Circos: An information aesthetic for comparative genomics. *Genome Res.* <https://doi.org/10.1101/gr.092759.109> (2009).

Acknowledgements

The author acknowledges the Department of Biotechnology (DBT) for granting fellowship. We would also like to thank the Director of the ICAR-National Dairy Research Institute (NDRI).

Author contributions

R.K.: Sample collection and preparation, formal analysis, devised methodology, investigation, data visualization, result interpretation, writing-original draft, writing- review & editing; S.A.A.: Data acquisition, ms/ms data analysis, data visualization, result interpretation, review & editing. S.K.S.: - assisted in sample collection and preparation, writing- review & editing V.B.: methodology, writing- review & editing; S.K.: Conceptualization, supervision, validation, writing- review & editing, proofread and approved the final version of the manuscript; A.K.M. & J.K.K.: all authors read and approved the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-91684-4>.

Correspondence and requests for materials should be addressed to S.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021