Post translational modifications of milk proteins in geographically diverse goat breeds

P. K. Rout & M. Verma

Goat milk is a source of nutrition in difficult areas and has lesser allergenicity than cow milk. It is leading in the area for nutraceutical formulation and drug development using goat mammary gland as a bioreactor. Post translational modifications of a protein regulate protein function, biological activity, stabilization and interactions. The protein variants of goat milk from 10 breeds were studied for the post translational modifications by combining highly sensitive 2DE and Q-Exactive LC-MS/MS. Here we observed high levels of post translational modifications in 201 peptides of 120 goat milk proteins. The phosphosites observed for CSN2, CSN1S1, CSN1S2, CSN3 were 11P, 13P, 17P and 6P, respectively in 105 casein phosphopeptides. Whey proteins BLG and LALBA showed 19 and 4 phosphosites respectively. Post translational modification was observed in 45 low abundant non-casein milk proteins mainly associated with signal transduction, immune system, developmental biology and metabolism pathways. Pasp is reported for the first time in 47 sites. The rare conserved peptide sequence of (SSSEE) was observed in αS1 and αS2 casein. The functional roles of identified phosphopeptides included anti-microbial, DPP-IV inhibitory, anti-inflammatory and ACE inhibitory. This is first report from tropics, investigating post translational modifications in casein and non-casein goat milk proteins and studies their interactions.

Milk is the primary source of nutrition for mammals and serves as a major vehicle of maternal immunity transfer, thus, plays a vital role in inclusive development of the neonates. The milk proteome is extremely complex due to abundant post-translational modifications and various proteolytic processes. Milk protein composition exhibited high heterogeneity due to numerous genetic variants, and isoforms with different degrees of posttranslational modifications such as phosphorylation and glycosylation in caseins. Milk proteins exhibit conformational structure due to post translational modifications and constitutive levels of proteolytic activity produce a range of significant peptides. The posttranslational modifications of the polypeptide chain occur in the Golgi apparatus of the mammary epithelial cells. Casein phosphorylation at amino acid serine or threonine is catalyzed by kinase enzymes. Phosphorylation is affected by different factors such as protein sequence, efficacy of kinase enzymes, gene expression, substrate availability and access to phosphorylation site which is responsible for the specific protein conformation.

Protein functions such as binding, stabilization, biological activity, interactions with proteins and other biomolecules are regulated by phosphorylation–dephosphorylation of protein. Phosphorylation stabilizes calcium phosphate nano clusters in casein micelles. The micellar structure of casein enables milk to carry calcium and phosphate to the neonate by channelizing the risk of mammary gland bio-calcification. Phosphorylation state of caseins varies widely from 1P to 3P on CSN3, 4P to 5P on CSN2, 8P to 9P on CSN1S1, 10P to 13P on CSN1S2. In bovine milk, CSN1S1 accounts for about 35% of the total casein and has 2 common phosphorylation isoforms: CSN1S1-8P and CSN1S1-9P. Similarly, CSN1S2 accounts for about 10% of the total casein and is present with isoforms from 10 to 14P and occasionally with 9P or 15P.

The identification and analysis of phosphopeptides has been challenging because of the relatively low stoichiometry, inherent lower ionization efficiency and variation of phosphorylation sites. However, advances in proteomics have largely enhanced the quotidian of protein identification. Phosphorylation of milk proteins has been studied in llama, camel and goat milk fat globule membrane proteins using MS/MS proteomic approach. A total of 8 phosphopeptides corresponding to 18 phosphorylation sites were identified in CSN1S1, CSN1S2 and CSN2 of goat milk using nLC-MS/MS. Casein phosphopeptides in goat milk have been studied by Olumee-shabon and Boehemer and phosphoproteome of goat milk fat globule membrane have been reported by Henry and others. Phosphorylation have been reported for bovine caseins using inductively coupled plasma

Genetics and Breeding Division, ICAR-Central Institute for Research on Goats, Makhdoom, Farah, Mathura, Uttar Pradesh 281122, India. Email: prashmirout@gmail.com
mass spectrometry (ICP-MS)\(^1\), equine CSN1S1 and and CSN2 by nESI-MS/MS\(^{20,21}\), donkey CSN2 by MALDI-TOF and nESI-MS/MS\(^2\) and for CSN1S1 and CSN2 of water buffalo by MS\(^2\).

Bovine milk is extensively used, due to its high biological value and plasticity as it can be transformed to cheese and several other dairy products. The functional knowledge of casein and whey proteins has been identified for the presence of bioactive peptides\(^{24}\). These peptides show various biological activities and are released by proteolytic digestion of caseins and milk proteins in gut or during fermentation. This study was designed to gain an insight into molecular diversity of goat milk proteins and to identify various degree of PTM in casein and non-casein proteins. The characterization of high and low abundant milk proteins by gel based proteomic approach i.e. 2DE and nLC-MS/MS. The present study analysed post-translational modification such as phosphorylation, oxidation, and carbamidomethylation on casein and non-casein proteins in goat milk. As post-translational modifications (PTMs) are acting as the major means of intracellular communication, therefore, the interactions of various proteins involved in post translational modifications were analysed.

### Results

The present study has analyzed the post translational modifications in low and high abundant proteins in goat milk of genetically diverse goat breeds/genotypes reared in varied ecological and grazing condition in India. The present study generated a comprehensive profile of PTM in goat milk and identified PTM sites in relation to sample variations.

#### The goat milk proteome.

The milk samples from 10 Indian goat breeds were investigated regarding their genetic variation, proteome composition and post translational modifications. The selected populations were mapped in their home tract and reared in semi-intensive system and the animals were apparently healthy. Altogether 1240 milk samples were analyzed by means of 14% SDS-urea-PAGE and different allelic combinations of caseins were resolved to identify the protein variants. Further, the selected variants were analyzed by combining highly sensitive 2DE and Q-Exactive nLC-MS/MS using Thermo Fisher in-house reference database. Distinctive protein spots (n=144) selected from 21 variants in 2DE gels were processed by nLC-MS/MS. The MS data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD013593. The .raw files generated from nLC-MS/MS of goat milk proteins were analysed for Sequest search using Proteome Discoverer (v2.2). By comparing the databases, a milk proteome of 578 peptides for 348 genes (800 Uniprot accessions) were identified using reference database. The post translational modifications were identified in 201 peptide sequences for 86 proteins (120 Uniprot KB accessions).

#### Post translational modifications in goat milk proteins.

The annotations of peptides observing modifications including cysteine carbamidomethylation, oxidation, acetylation and phosphorylation in each protein are presented in Table S2. Post translational modifications were observed in a total of 201 peptide sequences corresponding to 120 Uniprot ID across the reference databases. PTMs were observed in high and low abundant milk proteins in each breed. High performance Q-Exactive LC-MS/MS analyses identified 287 sites of phosphorylation on 120 unique phosphoproteins. The Ser/Thr/Tyr/Asp ratio was 128:70:42:47, respectively. The phosphorylation at Asp residues is reported for the first time in goat milk proteins.

### Phosphorylation of high abundance proteins.

The observed phosphosites for the high abundance proteins in goat milk are presented in Table 1. In the casein fraction, β-casein (CSN2) showed a maximum of 17 phosphosites, 4D, 5T, 6S and 2Y; α-S2 casein (CSN1S2) showed 13 phosphosites (7S, 1D, 5T); α-S1 casein (CSN1S1) showed 11 phosphosites (5S, 3D, 1T, 2Y); and κ-casein (CSN3) showed 6 phosphosites (3S, 1D, 2T), respectively. Whey proteins β-lactoglobulin (BLG) (P02756) showed 19 phosphosites (7D, 3Y, 5T, 4S) and α-lactalbumin (LALBA) showed 4 phosphosites (3D, 1Y).

The phosphorylated peptides generated from these proteins showed different isoforms and are presented in Table 2. The phosphorylated peptide 76–93 in CSN1S1 showed 5P at S78–S82 and peptide 58–73 showed 1P or 3P resulting in isoforms 10P or 12P. Similarly, peptides from CSN1S2 showed variations leading to different isoforms. Peptides 19–40, 19–37 showed 3 variations in site of phosphorylation such as 1P S24 or 2P S24, S25 or 3P S23, S26, S25; Peptide 142–153 showed 1P S145 or 1P S147 or 3P S145 T146 S147 and peptides 154–165/166 showed 1P S159 or 2P S159 T160 resulting in 8P isoform. β-casein peptides showed variations 1P or 6P in peptide 129–147, 2P or 3P in peptide 148–184 and may result in 10P and 11P isoforms. The identified fragment of κ-casein showed 4P to 5P variation by the level of phosphorylation of peptides 23–40 at 2P or 3P and 52–65/66 at 1P.

### Table 1. Identified phosphosites for the major milk proteins in Capra hircus reference database.

| Protein name | Gene ID | Number of phosphosites identified | Identified phosphorylation sites |
|--------------|---------|----------------------------------|--------------------------------|
| α-S1 casein | CSN1S1  | 11                               | S27, D58, S61, S63, T64, D66, Y95, D100, Y106, S130, S138 |
| α-S2 casein | CSN1S2  | 13                               | S23, S24, S25, S145, T146, S147, S151, T154, D156, S159, T160, T161, T164 |
| β-casein    | CSN2    | 17                               | S50, T56, D58, D62, Y129, T135, S137, S139, T141, T143, D144, S157, T167, Y169, S181, D197, Y206 |
| κ-casein (Fragment) | CSN3 | 6                               | S24, Y16, S60, T64, T73, S95 |
| α-lactalbumin | LALBA  | 4                                | D82, D83, D116, Y122 |
| β-lactoglobulin | BLG   | 19                               | D29, T36, Y38, S39, S45, S48, D51, S54, Y60, T67, T94, D114, T115, D116, Y117, T143, D147, D155 |
| Protein name | Peptide no. | Peptide sequence* | Level of phosphorylation | Master protein accessions | Positions in master proteins | # Protein groups | Proteins identified | # PSMs | # Missed cleavages |
|--------------|------------|-------------------|--------------------------|--------------------------|-----------------------------|-----------------|--------------------|--------|-------------------|
| αS1-casein   | 1          | HPINIQQALEPVL-    | 1P                       | Q8MIH4                   | [19–37]                     | 1               | 8                  | 1      | 0                |
|              | 2          | DIGSEI-EDQAMEDAK  | 1P                       | Q8MIH4                   | [58–73]                     | 3               | 13                 | 2      | 0                |
|              | 3          | DIGSEI-EDQAMEDAK  | 3P                       | Q8MIH4                   | [58–73]                     | 2               | 13                 | 3      | 0                |
|              | 4          | AGSSSSSEIVFNP-    | 5P                       | A0A0P0EL46               | [76–93]                     | 1               | 2                  | 5      | 0                |
|              | 5          | YQPEDVPSEER       | 1P                       | Q8MIH4                   | [95–105]                    | 3               | 11                 | 9      | 1                |
|              | 6          | YGYLEQLLR         | 1P                       | Q8MIH4                   | [106–115]                   | 2               | 8                  | 2      | 0                |
|              | 7          | KYNVQPQLEIVPK-    | 1P                       | Q8MIH4                   | [118–139]                   | 3               | 5                  | 2      | 2                |
|              | 8          | YNVQPQLEIVPK-    | 1P                       | Q8MIH4                   | [119–139]                   | 3               | 5                  | 1      | 1                |
|              | 9          | SAEQILHSMK        | 1P                       | Q8MIH4                   | [129–138]                   | 1               | 7                  | 1      | 0                |
|              | 10         | SAEQILHSMK        | 1P                       | Q8MIH4                   | [130–139]                   | 1               | 7                  | 4      | 0                |
| αS2-casein   | 1          | MEHVSSSEEPIN-    | 2P                       | P33049                   | [19–37]                     | 1               | 1                  | 2      | 0                |
|              | 2          | MEHVSSSEEPIN-    | 1P                       | P33049                   | [19–40]                     | 1               | 1                  | 2      | 1                |
|              | 3          | MEHVSSSEEPIN-    | 3P                       | P33049                   | [19–40]                     | 1               | 1                  | 3      | 1                |
|              | 4          | EQLSTSFEENKK      | 1P                       | P33049                   | [142–153]                   | 1               | 2                  | 2      | 1                |
|              | 5          | EQLSTSFEENKK      | 1P                       | P33049                   | [142–153]                   | 1               | 2                  | 1      | 1                |
|              | 6          | EQLSTSFEENKK      | 3P                       | P33049                   | [142–153]                   | 1               | 2                  | 3      | 1                |
|              | 7          | TIDMESTEVTFTK     | 2P                       | P33049                   | [154–165]                   | 1               | 2                  | 2      | 0                |
|              | 8          | TIDMESTEVTFTK     | 1P                       | P33049                   | [154–165]                   | 1               | 2                  | 1      | 0                |
|              | 9          | TIDMESTEVTFTK     | 1P                       | P33049                   | [154–166]                   | 1               | 2                  | 2      | 1                |
|              | 10         | TIDMESTEVTFTK     | 2P                       | P33049                   | [154–166]                   | 1               | 2                  | 10     | 1                |
| β-casein     | 1          | IEKFQSEQQQT-      | 1P                       | Q95L76                   | [45–63]                     | 1               | 4                  | 1      | 1                |
|              | 2          | FQSEQQQT-        | 1P                       | Q95L76                   | [48–63]                     | 1               | 4                  | 4      | 0                |
|              | 3          | YPVEFTEESQLTLT-   | 1P                       | Q95L76                   | [129–147]                   | 1               | 8                  | 1      | 0                |
|              | 4          | YPVEFTEESQLTLT-   | 6P                       | Q95L76                   | [129–147]                   | 1               | 8                  | 9      | 0                |
|              | 5          | YPVEFTEESQLTLT-   | 6P                       | P33048                   | [129–147]                   | 2               | 8                  | 22     | 0                |
|              | 6          | LHPLPVLQSWM-      | 2P                       | P33048                   | [148–184]                   | 2               | 8                  | 6      | 0                |
|              | 7          | LHPLPVLQSWM-      | 2P                       | P33048                   | [148–184]                   | 2               | 8                  | 9      | 0                |
|              | 8          | LHPLPVLQSWM-      | 2P                       | Q95L76                   | [148–184]                   | 1               | 8                  | 3      | 0                |
|              | 9          | DMPQALFFYQEVNL-   | 1P                       | Q95L76                   | [197–215]                   | 1               | 4                  | 2      | 0                |
|              | 10         | DMPQALFFYQEVNL-   | 1P                       | P33048                   | [197–215]                   | 2               | 4                  | 1      | 0                |
| κ-casein     | 1          | SPAQTLQWQVLPNPT-  | 2P                       | B2Z896                   | [23–40]                     | 4               | 45                 | 3      | 0                |
|              | 2          | SPAQTLQWQVLPNPT-  | 3P                       | B2Z896                   | [23–40]                     | 6               | 45                 | 15     | 0                |
|              | 3          | HPHHSLGMAIPPK     | 1P                       | B2Z896                   | [52–65]                     | 4               | 48                 | 1      | 0                |
|              | 4          | HPHHSLGMAIPPK     | 1P                       | B2Z896                   | [52–66]                     | 4               | 48                 | 1      | 1                |
|              | 5          | YIPQVYSLR         | 1P                       | Q7YRX4                   | [16–25]                     | 2               | 26                 | 2      | 0                |

Continued
Table 2. Identification of phosphopeptides of caprine αS1, αS2, β, κ caseins and α-lactalbumin and β-lactoglobulin by nano-LC-MS/MS. *Phosphorylated amino acids are in bold and underlined.

| Protein name | Peptide no. | Peptide sequence* | Level of phosphorylation | Master protein accessions | Positions in master proteins | # Protein groups | Proteins identified | # PSMs | # Missed cleavages |
|--------------|-------------|-------------------|--------------------------|--------------------------|-----------------------------|-----------------|-------------------|--------|-------------------|
| **β-lactoglobulin** | | | | | | | | | |
| 1 | GLDQKVKAGTW | YSLAMAASDSSL-DAGASPLR | 2P | P02756 | [27–68] | 1 | 2 | 2 | 1 |
| 2 | VAGTVYSLLAMAA- | DSSSL-DAGASPLR | 6P | P02756 | [33–58] | 1 | 3 | 34 | 0 |
| 3 | VAGTVYSLLAMAA- | DSSSL-DAGASPLR | 5P | P02756 | [33–58] | 1 | 3 | 14 | 0 |
| 4 | VAGTVYSLLAMAA- | DSSSL-DAGASPLR | 4P | P02756 | [33–58] | 1 | 3 | 4 | 0 |
| 5 | VAGTVYSLLAMAA- | DSSSL-DAGASPLR | 3P | P02756 | [33–58] | 1 | 3 | 8 | 0 |
| 6 | VAGTVYSLLAMAA- | DSSSL-DAGASPLR | 4P | P02756 | [33–58] | 1 | 3 | 2 | 0 |
| 7 | VYYVEELKPTPEGN- | LEIILOQK | 2P | P02756 | [59–78] | 1 | 3 | 5 | 0 |
| 8 | VYYVEELKPTPEGN- | LEIILOQK | 1P | P02756 | [59–78] | 1 | 3 | 1 | 0 |
| 9 | IIAEKFDIPAVFK | 1P | P02756 | [89–101] | 2 | 3 | 7 | 2 |
| 10 | TQIVPaviFk | 1P | P02756 | [94–101] | 1 | 3 | 2 | 1 |
| 11 | VLYLIDYKK | 2P | P02756 | [110–119] | 1 | 3 | 2 | 1 |
| 12 | TPEVDKEALEK | 1P | P02756 | [143–153] | 1 | 3 | 2 | 1 |
| 13 | TPEVDKEALEK | 2P | P02756 | [143–153] | 1 | 3 | 4 | 1 |
| 14 | TPEVDKEALEKFDK | 2P | P02756 | [143–156] | 1 | 3 | 7 | 2 |
| 15 | TPEVDKEALEKFDK | 1P | P02756 | [143–156] | 1 | 3 | 3 | 2 |
| **α-lactalbumin** | | | | | | | | | |
| 1 | JWCKDPQNPHSR | 2P | B2YKX6 | [78–89] | 2 | 3 | 6 | 1 |
| 2 | JWCKDPQNPHSR | 1P | B2YKX6 | [78–89] | 3 | 3 | 5 | 1 |
| 3 | KIILERYGNYWLAHK | 1P | P00712 | [113–127] | 1 | 1 | 4 | 2 |
| 4 | HLIKRVGNYWLAHK | 1P | P00712 | [114–127] | 1 | 1 | 7 | 1 |
| 5 | VGNYWLAHK | 1P | P00712 | [118–127] | 1 | 1 | 3 | 0 |

Similarly, LALBA peptide 78–89 showed 2P and 1P variation at D82, D83, peptide 113–127, 114–127 showed 1P at D116 and peptide 118–127 showed 1P at Y122 resulting in one more possible isoform up to 4P. In BLG the highest level of variation in phosphorylation was observed in peptide 33–58 with 3P, 4P, 5P and 6P at sites T36, Y38, S39, S45, D46 and S48. Furthermore, the peptides 59–78, 143–153/156 showed 1P and 2P variations resulting in 13P isoform of BLG. These findings indicated the different levels of isoforms in phosphorylation of the caprine milk proteins.

**Phosphorylation of low abundance proteins.** The sensitivity for the identification of phosphorylation sites in low abundance proteins was enhanced by combining 2DE and nLC-MS/MS approach. Phosphorylation was identified in 45 low abundant proteins and presented in Table 3. CREB binding protein (5P), inter-alpha-trypsin inhibitor heavy chain H2 (2P), olfactory receptor ORS1D1 (3P), GYCAM1 (1P), putative transcription factor inhibitor heavy chain (1P) were identified. Antiviral interferon tau BB4 found only in goat milk was also detected with 7P level of phosphorylation.

**Isoform in different breeds.** The identified 123 common UniprotKB ID proteins with varying level of phosphorylation across the 10 breeds were observed by comparing databases. The phosphosites and phosphoproteins observed in each breed are presented in Table S3. CSN1S1 (Q8M1H4)9P (S27, D58, S61, S63, T64, D66, Y95, Y106, S130) exhibited 10P isoform in Barbari, CSN2 (Q95L76) showed 13P (S50, D62, Y129, T135, S137, T141, T143, D144, S157, S167, T169, S181, D197) in Jamunapari goat. Phosphorylation at (S151, D156) and T164 was observed in Himalayan local goat and Osmanabadi goat, respectively. Similarly, the low abundant proteins were identified in different breeds with varying sites of phosphorylation. Proteins Proteoglycan 4 (PRG4) showed 8P isoform in Barbari, PIK3C2A, CEP152 observed 7P each and SLC1A6 1P in Barbari. Interferon tau BB4 (7P), CREBBP (5P), Cationic trypsin (8P) isoforms were observed in Jamunapari.

**PDZD9 and SMC5 observed exhibited 5P and 7P in Osmanabadi goats. Emilin-2 (3P), FAM222B (4P), MnSOD and OVOL1 each with 2P isoforms were observed in Sirohi goats.**
Other post translational modifications. Other post translational modification such as oxidation and carboxymethylation were also observed and presented in Table 4. Caseins showed only oxidation and phosphorylation. Peptides of LALBA and BLG showed both oxidation and carboxymethylation other than phosphorylation. β-casein showed carboxymethylation on 7 sites (M117, M124, M159, M160, M171, M172, M186) whereas; α-S2 casein displayed 3 oxidation sites M42, M157, M206. PTM in proteins α-lactalbumin were observed as C80, C110, C130, C139 and C84, C178, M42, M163 on β-lactoglobulin. Low abundant proteins also exhibited oxidation and carboxymethylation such as interferon tau BB4 (C123, M127), laminin subunit LAMC2 (C493, C496, C514, C517, C519, C534), olfactory

Table 3. Phosphorylation in low abundance proteins of goat milk identified in the reference database with accession number and gene name.

| Accession number | Protein name | Identified phosphosites | Gene name |
|------------------|--------------|-------------------------|-----------|
| A5JS7            | Glycosylation-dependent cell adhesion molecule-1 | S87        | GLYCAM1   |
| G0Z386           | Insulin-like growth factor binding protein-3 (Fragment) | S10, S13  | IGFBP3    |
| B74L4            | Interferon tau BB4 | S96, S97, D101; T102; T103; D117; D118 | INTERFERONAB |
| Q7YS14           | LDHA protein (Fragment) | S8, D9    | LDHA      |
| D2KMI2           | NADH-ubiquinone oxidoreductase chain 5 | D297       | ND5       |
| A0A0C53B61       | NADH-ubiquinone oxidoreductase chain 6 | T95; Y105; Y106 | ND6 |
| G1DGB8           | Putative transcription factor Ovo-like 1 | T149, Y150 | OYL1      |
| C8KG56           | Speroxide dismutase (Fragment) | T134, S143 | MnSOD     |
| W3P129           | Family with sequence similarity 222 member B | S111      | FAM222B   |
| W3P2D4           | Neurofilament heavy | S62        | NEFH      |
| W3PF40           | HECT and RLD domain containing E3 ubiquitin protein ligase 4 | S95/96/104 | HERC4  |
| W3P9A9           | Ofactory receptor | T281; S282; Y292 | OR51D1 |
| W3PGE4           | ATP/GTP binding protein 1 | S780; S792 | AGTPBP1   |
| W3PHS2           | Lipocalin_cytosolic FA-bd_dom-containing protein | T34       | LCN2      |
| W3QOT0           | Uncharacterized protein | S123      |           |
| W3QV29           | Peptidase S1 domain-containing protein | S179      |           |
| W3QVQ4           | Voltage-dependent calcium channel gamma-5 subunit | T102      | CACNG5    |
| W3QW7K           | Uncharacterized protein | T299      | LOC10110727 |
| W3QVN4           | Succinyl-CoA-A3-ketoadipate A transferase | S507      | OXCT1     |
| A7MB45           | Acyl-CoA synthase short-chain family member 3, mitochondrial | S20; S24; S25 | ACS63  |
| E1BDU1           | Ofactory receptor | Y120      | OR2K2     |
| E1BDZ8           | Enhancer of polycomb+R63b homolog | T551      | EPC2      |
| F1MD32           | CREB binding protein | Y226; T228; S247; S252; T258 | CREBBP |
| F1MNW4           | Inter-alpha-trypsin inhibitor heavy chain H2 | S543; S549 | ITIH2     |
| F1MVCO           | Uncharacterized protein | T201      | CAD       |
| G2X0Y3           | Mediator of RNA polymerase II transcription subunit 14 | T1097; T1098; S1101 | MED14 |
| P07670           | Cationic trypsin | S125; S127; S134; T130; S173; S175; T180; S181; S215 | PRSS1     |
| Q1HQA8           | LysM and putative peptidoglycan-binding domain-containing protein 2 | S21       | LYSMD2    |
| O00443           | Phosphatidylinositol 4-phosphate 3-kinase C2 domain-containing subunit alpha | T297; Y298; S299; Y398; D497; S498; D952 | PIK3CA2A |
| O94986           | Centrosomal protein of 152 kDa | S329; T331; T332; S338; D345; S349; S351 | CEP152   |
| P3471            | Receptor-type tyrosine-protein phosphatase zeta | S1444; S1446; S1447; Y1448 | PTPRZ1   |
| P46664           | Excitatory amino acid transporter 4 | S562      | SLC1A6    |
| Q13753           | Laminin subunit gamma-2 | D498; Y500; D503; D524; S526; S528; D532 | LAMC2    |
| Q5IYW2           | Centrosomal protein of 78 kDa | S117; S119; S120 | CEPI78   |
| Q7ZSA7           | Protein Fam19A5 | T78; T79; D87 | FAM19A5  |
| Q8MXQ2           | PDZ domain-containing protein 9 | S235; S236; S238; S239; S241 | PDZD9    |
| Q8Y118           | Structural maintenance of chromosomes protein 5 | T3; S5; T8; S9; T10; S12; S16 | SMC5     |
| Q8N638           | Ankyrin repeat domain-containing protein 35 | S349; S355 | ANKR35    |
| Q8N9J5           | Protein phosphatase 1K, mitochondrial | T352; S355; S360; S362 | PPM1K    |
| Q8N3Z6           | Zinc finger CCHC domain-containing protein 7 | S411      | ZCCHC7    |
| Q8WUS8           | Protein Fam222B | S111; D115; D117; T119 | FAM222B  |
| Q92554           | Proteoglycan 4 | Y38; S39; D41; T43; D47; Y48; Y53; D59 | PRG4     |
| Q96SR6           | Zinc finger protein 382 | S333; T341; T346; T347 | ZNF382   |
| Q98XX0           | EMILIN-2 | D544; D551; D573 | EMILIN2  |
| Q9H251           | Cadherin-23 | T259; D262; D264 | CDH23    |

Table 3. Phosphorylation in low abundance proteins of goat milk identified in the reference database with accession number and gene name.
| Modifications (oxidation, carboxymethylation) | Accession number | Protein description                        | Species | Gene name |
|---------------------------------------------|-----------------|-------------------------------------------|---------|-----------|
| C110, C130, C139                            | P00712          | Alpha-lactalbumin                          | Capra hircus | LALBA     |
| C122, M127                                  | B75L4           | Interferon tau B4                          | Capra hircus | INTERFERONAB |
| C125                                        | Q6SAN9          | Fatty acid binding protein 3               | Capra hircus | FABP3     |
| C148, C151                                  | G1DG8           | Putative transcription factor Ovo-like 1   | Capra hircus | OVO1      |
| C279; C291; M227                            | D2KMJ2          | NADH-ubiquinone oxidoreductase chain 5     | Capra hircus | ND5       |
| C42, M60, M97                               | B2Z986          | Kappa casein                               | Capra hircus | CSN3      |
| C494, M84/86                                | C6KJ77          | Breast cancer resistance protein           | Capra hircus | ABCG2     |
| C80                                         | A0A5M4RFP90     | Chemokine receptor 2                       | Capra hircus | CCR2      |
| C80, C110                                   | B2YKX6          | Alpha-lactalbumin                          | Capra hircus | LALBA     |
| C84; C178; M42; M163                        | P02756          | Beta-lactoglobulin                         | Capra hircus | LGB       |
| C9                                          | H2ETE9          | Natriuretic peptide C                      | Capra hircus | NPPC      |
| M11                                         | G0Z386          | Insulin-like growth factor binding protein-3 (Fragment) | Capra hircus | IGBF3     |
| M11, M124, M159, M171, M198                 | Q9SL76          | Beta-casein                                | Capra hircus | CSN2      |
| M117, M124, M198, M159, M160, M171, M172    | P33048          | Beta-casein                                | Capra hircus | CSN2      |
| M150, M138                                  | Q8MHJ4          | Alpha s1 casein                            | Capra hircus | CSN1S1    |
| M276                                        | A9UFM4          | Hormone sensitive lipase (Fragment)        | Capra hircus | HSL       |
| M33                                         | C6ZP43          | I alpha globin                             | Capra hircus | HBA1      |
| M36, M48; M75                               | Q5YD57          | Beta-casein                                | Capra hircus | CSN2      |
| M42, M157; M206                             | P33049          | Alpha-S2-casein                            | Capra hircus | CSN1S2    |
| M54, M70                                    | H1X3V0          | Promyelocytic leukemia zinc finger protein (Fragment) | Capra hircus | PLZF      |
| M60, M63                                    | G1DFR8          | Histone H2B type 3-A                       | Capra hircus | HIST3H2BA |
| M88                                         | Q6R649          | Keratin, type I cytoskeletal 27            | Capra hircus | KRT27     |
| M90                                         | Q7YRX4          | Kappa-casein (Fragment)                    | Capra hircus | CSN3      |
| M99, M103                                   | A0A0C5B361      | NADH-ubiquinone oxidoreductase chain 6     | Capra hircus | ND6       |
| C186, C189                                  | Q0ZBS4          | Pol protein (Fragment)                     | Caprina arthritis encephalitis virus | POL       |
| M459                                        | Q9DVK8         | Pol protein                               | Caprina arthritis encephalitis virus | POL       |
| C111                                        | E1BGN3          | Histone H3                                | Bos taurus | HIST2H3D  |
| C125                                        | P10790          | Fatty acid-binding protein, heart          | Bos taurus | FABP3     |
| C156                                        | F1MCF8          | Ig-like domain-containing protein          | Bos taurus | HAVCR2    |
| C176; M161; M48; C82; C122; C135; C137;     | G5S5H7          | Deleted entry                             | Bos taurus | LOC615237 |
| C183                                        | Q7YS80          | Myogenic factor 6                         | Bos taurus | MYP6      |
| C30; C48; C64; C132; C139; C160; C171; C185; | P00760          | Cationic trypsin                          | Bos taurus | PRS1      |
| C196; C206; C220; C233; M109; M183           | C47, C130; C139 | Alpha-lactalbumin                          | Bos taurus | LALBA     |
| C97, C112; M101; M107; M118                  | E1BDU1          | Olfactory receptor                         | Bos taurus | OR2K2     |
| M124; M200                                  | P02666          | Beta-casein                               | Bos taurus | CSN2      |
| M125                                        | F1MC11          | Keratin, type I cytoskeletal 14            | Bos taurus | KRT14     |
| M127                                        | P02668          | Kappa-casein                              | Bos taurus | CSN3      |
| M196                                        | F1MHA3          | Ig-like domain-containing protein          | Bos taurus | TARP      |
| M562                                        | F1MNW4          | Inter-alpha-trypsin inhibitor heavy chain H2 | Bos taurus | ITIH2     |
| M60; M63                                    | E1BRK9         | Histone H2B                               | Bos taurus | HIST3H2BB |
| M93                                         | F1MQI3          | Peptidase S1 domain-containing protein     | Bos taurus | LOC615237 |
| Acetylation                                  | Q8IY18        | Structural maintenance of chromosomes protein 5 | Homo sapiens | SMCC5     |
| C111                                        | Q7ID13          | Histone H3.2                              | Homo sapiens | HIST2H3A  |
| C115                                        | Q5ITW2          | Centrosomal protein of 78 kDa             | Homo sapiens | CEPT8     |
| C1442; C1445                               | P23471          | Receptor-type tyrosine-protein phosphatase zeta | Homo sapiens | PTTPRZ1    |
| C249                                        | Q8XXQ8          | PDZ domain-containing protein 9            | Homo sapiens | PDZD9     |
| C49; C265; M262; M296                       | P04264          | Keratin, type II cytoskeletal 1            | Homo sapiens | KRT1      |
| C493; C496; C514; C517; C519; C531           | Q13753          | Laminin subunit gamma-2                   | Homo sapiens | LAMC2     |
| C553; M564                                  | Q9BXE0          | EMILIN-2                                  | Homo sapiens | EMILIN2   |
| C77                                         | P02538          | Keratin, type II cytoskeletal 6A           | Homo sapiens | KRT6A     |
| C85                                         | Q7Z5A7          | Protein FAM19A5                            | Homo sapiens | FAM19A5   |

Continued
### Table 4. Oxidation and carboxymethylation sites as observed in high and low abundance goat milk proteins.

| Modifications (oxidation, carboxymethylation) | Accession number | Protein description | Species | Gene name |
|-----------------------------------------------|------------------|---------------------|---------|-----------|
| M109                                          | P07477           | Trypsin-1           | Homo sapiens | PRSS1     |
| M119                                          | P02533           | Keratin, type I cytoskeletal 14 | Homo sapiens | KRT14     |
| M150; M271; M306                              | P13645           | Keratin, type I cytoskeletal 10 | Homo sapiens | KRT10     |
| M157; M234; M245; M324; C406                  | P35527           | Keratin, type I cytoskeletal 9 | Homo sapiens | KRT9      |
| M16                                           | Q8NCf6           | Beta-galactosidase-1-like protein 3 | Homo sapiens | GLR1L3    |
| M2243                                         | Q96RW7           | Hemicentin-1        | Homo sapiens | HMCN1     |
| M334                                          | O94986           | Centrosomal protein of 152 kDa | Homo sapiens | CEP152    |
| M406                                          | Q13873           | Bone morphogenetic protein receptor type-2 | Homo sapiens | BMP2      |
| M66; M63                                      | O68014           | Histone H2B type 1-K | Homo sapiens | HIST1H2BK |
| C66, C72                                      | F1AWZ7           | PP107               | Orf virus  |           |
| Acetylation                                    | W5SQP7           | Collin-1            | Orvis aries | CFL1      |
| C115                                          | W5SP91           | Histone H3          | Orvis aries | H3FA      |
| C30; C39; C47; C80; C110; C130; C139          | W5QDS2           | Alpha-lactalbumin   | Orvis aries | LALBA     |
| C708                                          | W5P440           | HECT and RLD domain containing E3 ubiquitin protein ligase 4 | Orvis aries | HERC4     |
| C811; M790                                    | W3PGE4           | ATP/GTP binding protein 1 | Orvis aries | AGTBPB1   |
| M102                                          | W5PXK1           | Peptidase S1 domain-containing protein | Orvis aries | PRSS1     |
| M102; M15; M189; C36; C45; C70; C177; C191   | W5Q219           | Peptidase S1 domain-containing protein | Orvis aries | PRSS1     |
| M124; M159; M198                              | P11839           | Beta-casein         | Orvis aries | CSN2      |
| M125; M199                                    | W5PLC2           | Beta-casein         | Orvis aries | CSN2      |
| M127                                          | A0A5979Y5N6      | Kappa-casein        | Orvis aries | CSN3      |
| M138; M150                                    | D2D3J8           | Alpha s1 casein     | Orvis aries | CSN1S1    |
| M157; M206                                    | E7BQS1           | Alpha-S2-casein     | Orvis aries | CSN1S2    |
| M163; C84; C178                               | P67976           | Beta-lactoglobulin  | Orvis aries | BLG       |
| M22                                           | R4R2H5           | Beta-casein         | Orvis aries | CSN2      |
| M288                                          | W3P9A9           | Olfactory receptor  | Orvis aries | ORS1D1    |
| M408                                          | W5Q885           | Receptor protein serine/threonine kinase | Orvis aries | BPM2      |
| M47; M48; M130; M139                          | P09462           | Alpha-lactalbumin   | Orvis aries | LALBA     |
| M518; C502                                    | W5QF4N4          | Succinyl-CoA:3-ketoacid-coenzyme A transferase | Orvis aries | OXCT1     |
| M60; M63                                      | W5Q2A4           | Histone H2B         | Orvis aries | HIST2H2BE |
| M84                                           | W5Q6L8           | IF rod domain-containing protein | Orvis aries | VIM       |
| M479                                          | A0A688DBX4       | Haemagglutinin protein | Peste-des-petits-ruminants virus | H         |

**Functional prediction and protein interaction analysis and bioactive peptides.** The identified proteome was classified into functional categories such as biological process (BP) (87.8% genes), cellular component (CC) (97.6% genes) and molecular function (MF) (92.7% genes) using DAVID 6.8. The detailed annotations including UP-Keywords from DAVID software are available in Table S4. Enriched GO terms for BP included single-multicellular organisomal process (P value 3.63 E−08), cellular component organization (P value 9.51 E−10), response to stress (P value 1.27 E−07), anatomical structure development (P value 2.94 E−07), defense process (P value 4.99 E−07), response to hypoxia (P value 2.57 E−04). The proteins mostly localised in membrane bound organelle (P value 1.72 E−06), and extracellular region (P value 4.05 E−46) and extracellular exosome (P value 1.27 E−51). The molecular functions (MF) largely involved binding activity such as protein binding, DNA binding, nucleosome binding, anion binding, antioxidant activity (P value 2.94 E−04) and structural molecular activities (P value 2.51 E−39) were also observed. The UP_Keywords comprised antimicrobials, phosphoprotein, disease mutation and methylation.

The interaction of proteins was analysed by STRING by MCL clustering of the identified proteome categorised in 16 clusters at highest confidence (>90%) score as depicted in Figure 1a. The network comprised of three major groups where keratin proteins shared distinct close interconnected cluster. Proteins found in defense response were also involved in response to stress, signal transduction and tissue development.

**Functional analysis of PTM proteins.** The GO functional annotations for PTM genes subset were enriched using DAVID 6.8 for classification into BP (87.5% genes), CC (94.4% genes) and MF (87.5% genes). The details of annotations are presented in Table S5. Enriched GO terms included developmental process (P value 0.02069), response to stress (P value 0.0099), response to drug (P value 0.0040), biogenesis (P value 0.048), structural...
Figure 1. Functional prediction and protein-protein interaction analysis: MCL clustering by STRING for all the identified proteins (a); Cytoscape network analysis of PTM gene subset (b); MCL clustering by STRING for PTM protein subset (c); Signal transduction pathway from reactome database for PTM protein subset (d); functional attributes of identified bioactive peptides (e).
molecular activity (P value 5.19 E−04), protein dimerization activity (P value 0.0092) and extracellular space (P value 3.00 E−06).

The phosphoprotein subset when analysed in Cytoscape Cluego identified 76 genes to generate a network with 6 functional groups at different levels. The details of annotations are presented in Table S6, Supplementary data file 1. The categorized functional groups were lactation, response to hydrogen peroxide, positive regulation of ossification, defense response to gram-positive bacterium, cornification and Systemic lupus erythematosus. Keratin proteins were associated with cornification, H2BC12, H2BC21 and KRT6A with antimicrobial humoral immune response mediated by antimicrobial peptide. Proteins BMPR2, NPPC, ZBTB16 found in defense response to gram-positive bacteria, CAD, CSN2, CSN3 in lactation and neurofilament in structural constituent of cytoskeleton. Estrogen signaling pathway was detected and proteins H3C1, H3C13, H3C15 were associated with pathways such as alcoholism, histone modifications, systemic lupus erythematosus (Fig. 1b).

The PTM gene subset exhibited a network comprising of 8 clusters in STRING (Fig. 1c) The keratin proteins clustered with PRSS1 (Trypsin); CSN2, CSN1S1, CSN3 and LALBA were identified in single cluster and histone proteins interacted closely with CREB-binding protein.

The identifiers (PTM genes) were analysed in Reactome database, where 61 identifiers were found in 332 pathways hit by at least one of them. The majority of proteins were associated with signal transduction, immune system, developmental biology and metabolism pathways. Pathways such as signaling by nuclear receptors, WNT, Notch and RHO GTPase were significant with p < 0.05 (Fig. 1d). The identified 17 phosphoproteins involved in signal transduction pathways included BMPR2, CCR2, CFL1, CREBBP; CSN2, H3FA, HIST1H2BK, HIST2H2BE, HIST2H3A, HIST2H3D, HIST3H2BB, LAMC2, OR2K2, ORS1D1 and PRG4. The details of pathways has been given in Table S7, Supplementary data file 1.

**Identification of bioactive peptides.** The 201 peptide sequences with identified PTM sites were grouped as long (>25 amino acids), medium (7–25 amino acids) and small (<7 amino acids). The identified phosphopeptides were categorized as long (62 peptides), medium (145 peptides), and small (1 peptide). The functions of these peptides were determined from MBPDB for 80% alignment with the known peptides are presented in Table 5. αS1-casein peptide 106–115 was anxiolytic and peptide 107–116 was antimicrobial. κ-casein peptide 46–55

| Search peptide | Protein ID | Peptide | Protein description | Position in protein | Function | % alignment | e-value | Alignment length | mismatches |
|----------------|------------|---------|---------------------|--------------------|----------|-------------|--------|-----------------|------------|
| ALKALPMHIRM   | P02754     | ALKALPMHIRM β-lactoglobulin | 155–164           | stimulates prolif-  eration | 100      | 1.63E–10   | 10     | 0               |
| ALPMHIRM      | P02754     | ALPMHIRM β-lactoglobulin | 158–164           | stimulates prolif- eration, Reduced vasconstrictoren- dothelin-1 release, ACE-inhibitory | 100      | 8.50 E–07  | 7      | 0               |
| AMKPWIQPK     | P02663     | AMKPWIQPK α-S2-casein | 204–212          | ACE-inhibitory     | 88.89 | 7.75 E–07  | 9      | 1               |
| DMPQAF-LYQEPVLGPVR | P02666 | DMPQAF-LYQEPVLGPVR β-casein | 199–217          | Anti-inflamma- tory | 94.73 | 3.33E–19   | 19     | 1               |
| DMPQAF-LYQEPVLGPVR | P02666 | DMPQAF-LYQEPVLGPVR β-casein | 199–217          | Anti-inflamma- tory | 100 | 1.39E–21  | 19     | 0               |
| EMPFPK        | P02666     | EMPFPK β-casein | 123–128           | ACE-inhibitory, Increased MUC4 expression, Bradykinin-potentiating, Antimicrobial | 100      | 1.49 E–05  | 6      | 0               |
| HKEMPFPK      | P02666     | HKEMPFPK β-casein | 121–128          | Antimicrobial      | 100 | 4.88 E–08  | 8      | 0               |
| IAEKTIPAVFK   | P02576     | IAEKTIPAVFK β-lactoglobulin | 89–100          | Antimicrobial      | 92.30 | 8.33E–13   | 12     | 0               |
| TPEVDKEALEK   | P02574     | TPEVDKEALEK β-lactoglobulin | 141–151         | DPP-IV Inhibi- tory, Antimi- crobial | 90.91 | 2.50 E–09  | 11     | 1               |
| TPEVDKEALEK   | P02576     | TPEVDKEALEK β-lactoglobulin | 143–152         | Antimicrobial      | 90.91 | 2.49E–10   | 10     | 0               |
| VGINYWLHAK    | P00711     | VGINYWLHAK α-lactalbumin | 118–127         | ACE-inhibitory     | 100 | 1.63E–10  | 10     | 0               |
| VKETMVFK      | F33048     | KETMVFK β-casein | 114–120          | Antimicrobial      | 87.5  | 1.35 E–06  | 7      | 0               |
| VLVLDTDYKK    | P02574     | VLVLDTDYKK β-lactoglobulin | 108–116         | DPP-IV Inhibi- tory, Antimi- crobial | 90 | 4.34 E–09  | 9      | 0               |
| VYVEELKPT-PEGDELIILQK | P02574 | VYVEELKPT-PEGDELIILQK β-lactoglobulin | 57–76         | Hypocholester- olemic | 95 | 1.97E–20  | 20     | 1               |
| YIPQYVLSR    | P02668     | YIPQYVLSR κ-casein | 46–55            | C3a Receptors agonist, Opioid | 100 | 1.63E–10  | 10     | 0               |
| YLGYEQLLR    | P02662     | YLGYEQLLR α-S1-casein | 106–115         | Anxiolytic         | 100 | 1.63E–10  | 10     | 0               |
| YLGYEQLLR    | P02662     | YLGYEQLLR α-S1-casein | 107–116         | Antimicrobial      | 90 | 6.03 E–09  | 9      | 0               |

**Table 5.** The function of identified bioactive peptide from MBPD.
The identified PTM sites varied in milk samples of different goat breeds (Table 4). Identification of low abundance phosphopeptides assigned to these proteins were mainly mono- or bi-phosphorylated (Table 2). To overcome the limited entries in the caprine database, other reference database were used for identification of present study and confirming earlier reports.

The present study focused on obtaining a comprehensive profile of the PTM sites of goat milk casein and non-casein protein and their interaction. Proteomics as a tool have been employed for the discovery, and characterization of post translational modifications such as phosphorylation, oxidation30–32. Electrospray ionization (ESI) mass spectrometry (MS) is suitable for studying PTM, including phosphorylation and glycosylation, since the technique provides native molecular mass determination of proteins. In this study, we analysed the goat milk proteins to identify both casein and non-casein PTM sites using nLC-MS/MS. We processed distinct protein spots and protein–protein interaction. Proteomics as a tool have been employed for the discovery, and characterization of post translational modifications such as phosphorylation, oxidation30–32. Electrospray ionization (ESI) mass spectrometry (MS) is suitable for studying PTM, including phosphorylation and glycosylation, since the technique provides native molecular mass determination of proteins. In this study, we analysed the goat milk proteins to identify both casein and non-casein PTM sites using nLC-MS/MS. We processed distinct protein spots and protein–protein interaction.

The present study focused on obtaining a comprehensive profile of the PTM sites of goat milk casein and non-casein protein and their interaction. Proteomics as a tool have been employed for the discovery, and characterization of post translational modifications such as phosphorylation, oxidation30–32. Electrospray ionization (ESI) mass spectrometry (MS) is suitable for studying PTM, including phosphorylation and glycosylation, since the technique provides native molecular mass determination of proteins. In this study, we analysed the goat milk proteins to identify both casein and non-casein PTM sites using nLC-MS/MS. We processed distinct protein spots and protein–protein interaction. Proteomics as a tool have been employed for the discovery, and characterization of post translational modifications such as phosphorylation, oxidation30–32. Electrospray ionization (ESI) mass spectrometry (MS) is suitable for studying PTM, including phosphorylation and glycosylation, since the technique provides native molecular mass determination of proteins. In this study, we analysed the goat milk proteins to identify both casein and non-casein PTM sites using nLC-MS/MS. We processed distinct protein spots and protein–protein interaction.

Phosphorylation has been well characterized in the bovine caseins33,34. There are various reports targeting casein fractions in bovine and non-bovine milk by various proteomic approaches32,35–39. The majority of bovine caseins exist in a phosphorylated form, and the phosphorylated residues vary from individual variants possessing one phosphorylated residue (κ-CN) to 13P for others (αs2 CN)40. Bilj and others41 demonstrated that high αs1-CN-8P concentration in bovine milk is a great benefit for the production of uncooked curd cheese because αs1-CN-8P is hydrolyzed more efficiently by chymosin during ripening. Bovine proteome analysis showed more than 30 phosphorylated proteins which included 5 CSN2, 15 CSN1S1, 10 CSN1S2 and 4 CSN3 casein components42. Similarly, donkey milk showed 11 CSN3, 6 CSN1S1 and 3 CSN1S2 casein components43.

Goat milk proteins have been analysed using MS/MS44–49. The present study resulted in 105 phosphopeptides from casein and non-casein proteins from all the analysed samples. The conserved peptide sequence of (SSSEE) in casein was also observed in CSN1S1 and CSN1S2 at 1P, 2P and 3P. The phosphosites were identified in casein and whey proteins and on 45 other low abundance proteins. The total number of phosphosites observed in the major milk proteins associated with S, T, Y and D residues were 32, 18, 11 and 21, respectively. The phosphorylation sites observed for CSN2, CSN1S1, CSN1S2, and CSN3 were 11P, 13P, 17P and 6P, respectively (Table 1). However, whey proteins BLG showed 19 phosphosites (7D, 3Y, 5T, 4S) and LALBA showed 4 phosphosites (3D, 1Y). The identified phosphopeptides resulted in 1P, 8P, 11P, 5P, 13P and 4P isoforms of CSN1S1, CSN1S2, CSN2, CSN3, BLG and LALBA, respectively. Beta casein showed highest degree of variation in phosphorylation with identification of 17P sites and other PTM such as oxidation in the identified peptides. A higher number of casein phosphopeptides and phosphorylation sites are reported in the present study in comparison to previous study.

Casein and whey proteins are post translationally modified by proteolysis by the milk enzymes, formation of disulphide bond by oxidation of cysteine, differential phosphorylation levels of serine and threonine, and glycosylation of threonine residues. The phosphorylation degree of αs-casein is a prime factor affecting the technological properties of milk. Therefore, “signature peptides” and “caseome” analysis are being used to investigate adulteration in milk of different species. The identification of cheese from different species has been authenticated by proteolytic peptides44,45. Therefore, proteome analysis of fermented milk products should be carried out due to their nutritional and health economic importance.

The present study reported 45 non casein phospho proteins assigned to various metabolic pathways (Table 3). The identified PTM sites varied in milk samples of different goat breeds (Table 4). Identification of low abundance proteins in milk is difficult as single step analysis fails to detect a large proportion of these proteins. Moreover, to overcome the limited entries in the caprine database, other reference database were used for identification of low abundance proteins. The varying levels and sites of phosphorylation in different breeds may be attributable to various physiological or environmental conditions under the influence of different agro-climatic regions. The phosphopeptides assigned to these proteins were mainly mono- or bi-phosphorylated (Table 2). The other post translational modifications such as acetylation, oxidation, and carbamidomethylation have also been reported (Table 4). The post translational modifications play an important role in protein/peptide...
functioning and their interaction. N-terminal acetylation increases peptide stability by preventing N-terminal degradation. Peptides with carbamidomethylation are mainly used in peptide mass fingerprinting for identification and characterization of proteins.

In the present study, the protein-protein interaction was analysed to know about the functional properties of proteins. The GO annotations were analysed using DAVID, network and interactions using Cytoscape and STRING and pathways analysis using Reactome database. The keratin proteins interacted with trypsin (PRSS1) and histone proteins interacted with CREB binding protein (CREBBP). The identified phosphoproteins were associated with lactation, response to stress, histone modification, cornification and signaling pathways (Tables S5, S6, S7). It has been reported that caseins are associated with other secreted calcium (phosphate)–binding phosphoproteins, such as osteopontin, in milk. Protein phosphorylation is vital for the regulation of metabolism, proliferation, inflammation, apoptosis, signaling and other important physiological processes. Autophosphorylation increases the catalytic efficiency of the receptor and provides binding sites for the assembly of downstream signaling complexes. Caseins form micelles which vary from species to species, and when cleaved, generated bioactive peptides, having potential functions making them protein of interest. This gastrointestinal degradation may be the consequence of enzymatic hydrolysis, fermentation and other processes used in dairy production.

The identification and characterization of phosphorylation sites are required to explore signaling networks of milk proteins. Phosphorylation-site provides definitive information on functional relationships between signaling proteins. The peptides released by enzymatic hydrolysis have specific biological functions due to their functional and interactions at cellular level. The identified phospho bioactive peptides were mainly anti-microbial followed by ACE inhibitory, DPP-IV inhibitory, proliferating functions. Anti-oxidative, anti-inflammatory and hypcholesterolemic peptides were also confirmed from goat milk proteins. The peptides released in the gastrointestinal degradation may be the consequence of enzymatic hydrolysis, fermentation and other processes used in dairy production.

Conclusion
Proteomic analysis has been carried out to study human, bovine and non-bovine milk for nutrition and therapeutic applications. Phosphorylation has been well characterized due to several technical, practical and bioinformatics approach. The present study identified 201 peptides showing post translational modifications in goat milk. The phosphorylation at Asp residues is reported first time in goat milk proteins. The rare conserved peptide sequence (SSSEE) in casein was observed in casein phosphopeptides (CSN1S1, CSN1S2). Phosphorylation regulates protein functions, such as biological activity, interaction and stabilization by causing conformational changes in the protein. Therefore, the identification of the post-translational modifications of the milk proteome has become necessary and may provide newer insights to extend the milk proteome and its potential biological role.

Materials and methods
The work flow of the present study has been depicted in Fig. 2.

Milk collection and description of genetic stock. Goat milk samples were collected from 1240 animals belonging to 10 goat breeds/genotypes during the postpartum days 30–65. Milk samples were collected from natural habitats of the breeds belonging to different geographical and agro-climatic regions of India. Samples were collected between 7:30 and 9:00 hrs. by hand milking. After disinfecting the udder, 30–40 ml of milk was collected directly to the collection tubes, transported to the laboratory at 4ºC within 24 hrs. and stored at -20ºC. Milk subsamples for protein analysis were stored at -40ºC until further analysis. All sample collection was conducted in accordance with institutional practice and the study was approved by Institutional animal ethics committee (IAEC).

The breeds analysed in the study belonged to arid, semi-arid, humid, coastal and mountain regions with different grazing conditions. The details of samples collected with the description of natural habitats from each breed are presented in Table S1, Supplementary data file 1. The animals were apparently healthy and the body condition score was satisfactory (3–4). The animals are mainly reared under semi-intensive system depending mostly on field grazing and supplementation of dry fodder, concentrate and mineral mixture.

Gel based milk protein analysis. The milk protein variants were analysed by SDS-PAGE and details have been described elsewhere. The skimmed milk samples were centrifuged at 12000g, 4ºC, 15 min to obtain clear transparent aqueous layer. This layer was separated, quantified and reduced in Laemmli’s sample buffer for analysis by SDS-PAGE. Milk proteins were resolved on 14% SDS and urea PAGE using 8 µg in each lane. The gels were stained with comassie brilliant blue (R250) and subsequently scanned in the Gel documentation system (Alpha Innotech Corporation, USA). Milk protein variants were identified by comparing the allelic variation with reference samples (confirmed by sequencing) by determining the molecular weight.

The protein variants (n=21) identified in the 10 breeds using SDS-PAGE were selected for analysis by 2DE. Individual protein samples were subjected to in–gel rehydration of IPG strips (SERA IPG Bluestrip 3–10 NL/7cm). Rehydration was carried out at room temperature for 16 hrs. with protein sample diluted in rehydration buffer.
Figure 2. Flow diagram showing the methodology of PTM analysis.

Figure 3. Geographical location of goat milk sampling. The breeds are abbreviated in the region of geographical origin. GD, Gaddi; HL, Himalayan local; JK, Jakhrana; B, Barbari; J, Jamunapari; S, Sirohi; BB, Black Bengal; GN, Ganjam; OS, Osmanabadi; AT, Attapady Black.
buffer (8M Urea, 0.002% bromophenol blue, 2% w/v CHAPS, 3 mg dithiothreitol, 1.5% ampholyte pH 3–10). For first dimensional electrophoresis, IPG strips were transferred to Hoefer IPG phor II. Iso electro focusing (IEF) was then performed at 20°C by a series of steps as follows: constant 250 V, 1:00 hrs.; constant 500 V, 1:00 hrs.; gradient 1000 V, 1:00 hrs.; gradient 3000 V, 2:00 hrs.; constant 3003 V, 2:00 hrs. The strips were then equilibrated in SDS equilibration buffer (8M Urea, 2% Tris HCl, 0.002% Bromophenol Blue) for 20 min. and loaded onto a 14% acrylamide gel for second dimension resolution. Second dimension run was carried out with mini gel SE 260 at 4V/cm. The 2DE protein spots were visualised by staining with 0.2% silver nitrate solution.

Nano-liquid chromatography mass spectrometry (nLC-MS/MS). The details of nLC-MS/MS analysis has been described elsewhere and used with a minor modifications for PTM analysis.

a. Protein digestion

The silver stained protein spots in gels were excised carefully with help of a sterilized spatula and transferred into separate vials. Destaining and acetone precipitation was carried out. In-gel digestion was performed; samples were reduced with 5 mM TCEP (Tris 2-Carboxyethyl Phosphine) at 55°C for 1 hour and further alkylated with 50 mM iodoacetamide for 30 min at room temperature in the dark. The gel spots were shrunked with acetonitrile and air-dried for few minutes at room temperature followed by digestion with trypsin (1:50, trypsin/lysate ratio) for 16 hours at 37°C. Digests were dried using speed vac for 1 hour and pellet was dissolved in buffer A (5% acetonitrile, 0.1% formic acid). Digests were cleaned by Sep-Pak. Titanium method was performed for phospho binding. Digests were cleaned up again using C18 silica cartridge (The Nest Group, Southborough, MA) following manufacturer’s protocol and dried using speed vac. The desalted dried pellet was reconstituted in buffer A (5% acetonitrile, 0.1% formic acid).

b. Liquid chromatography mass spectrometry analysis

All the experiments were performed using EASY-nLC 1000 system (Thermo Fisher Scientific) coupled to Q Exactive mass spectrometer (Thermo Fisher Scientific, Germany) equipped with nano electrospray ion source. Peptide mixture (1.0 µg) was loaded on a precolumn and was resolved using 5 cm PicoFrit column (360 µm outer diameter, 75 µm inner diameter, 10 µm tip) filled with 1.9 µm of C18-resin (Dr Maeisch, Germany). The peptides were loaded with buffer A and eluted with a 0–40% gradient of buffer B (95% acetonitrile, 0.1% formic acid) at a flow rate of 500 nl/min for 10 min. The QExactive was operated using the Top10 HCD data-dependent acquisition mode with a full scan resolution of 70,000 at m/z 400. MS/MS scans were acquired at a resolution of 17500 at m/z 400. Lock mass option was enabled for polydimethylcyclosiloxane (PCM) ions (m/z = 445.120025) for internal recalibration during the run. MS/MS data was acquired using a data-dependent top 10 method dynamically choosing the most abundant precursor ions from the survey scan.

c. Protein identification and PTM analysis

The .raw files generated were analyzed using Proteome Discoverer (v2.2) against the in-house Uniprot reference proteome database (Capra hircus, Ovis aries, Homo sapiens and Bos taurus). Due to scarcity of available protein annotations for goat, the analysis was carried out with goat, human, sheep and cow database with 7056, 27666, and 23869 entries respectively. For Sequest search, the precursor and fragment mass tolerances were set at 10–15 ppm and 0.5 Da, respectively. The protease used to generate peptides, that is the enzyme specificity was set for trypsin/P (cleavage at the C terminus of “K/R” unless followed by “P”) along with maximum missed cleavage value of 2. Carbamidomethyl on cysteine as fixed modification and oxidation of methionine and fucity was set for trypsin/P (cleavage at the C terminus of “K/R”: unless followed by “P”) along with maximum missed cleavage value of 2. Carbamidomethyl on cysteine as fixed modification and oxidation of methionine and fucity was set for trypsin/P (cleavage at the C terminus of “K/R”: unless followed by “P”) along with maximum missed cleavage value of 2. Carbamidomethyl on cysteine as fixed modification and oxidation of methionine and
References

1. D’Ambrosio, C. et al. A proteomic characterization of water buffalo milk fractions—describing PTM of major species and the identification of minor components involved in nutrient delivery and defense against pathogens. Proteomics 8, 3657–3666 (2008).

2. Holland, J. W. & Boland, M. J. Post-translational Modifications of Caseins. 2nd ed. Singh, H., Thompson, A. & Boland, M. (Eds.) Elsevier Inc., London, UK (2014).

3. Caroli, A. M., Chessa, S. & Erhardt, G. J. Invited review: milk protein polymorphisms in cattle: effect on animal breeding and human nutrition. J. Dairy Sci. 92(11), 5333–52 (2009).

4. Bingham, E. W. & Farell, H. M. Phosphorylation of casein by the lactating mammary gland: a review. J. Dairy Sci 60(8), 1199–1207 (1997).

5. Mercier, J. C. Phosphorylation of caseins, present evidence for an amino acid triplet code posttranslationally recognized by specific kinases. Biochimie 63, 1–17 (1981).

6. Holland, J. W. Post-translational modifications of caseins. In Milk Proteins: From Expression to Food (eds Thompson, A. et al.) 107–132 (Elsevier, Amsterdam, 2009).

7. Cohen, P. The origins of protein phosphorylation. Nat. Cell Biol. 4, E127–E130 (2002).

8. Huppertz, T. Chemistry of the caseins. In Advanced Dairy Chemistry. Vol. 1A: Proteins: Basic Aspects (eds McSweeney, P. L. H. & Fox, P. E.) 135–160 (Springer, New York, NY, 2013).

9. Neville, M. C. Calcium secretion into milk. J. Mammary Gland Biol Neoplasia 10, 119–128 (2005).

10. Holt, C. & Carver, J. A. Darwinian transformation of a’ scarcely nutritious fluid’ into milk. J. Dairy Sci. 92(8), 3640–3615 (2009).

11. Farrell, H. M. Jr. The origins of protein phosphorylation. Mol. Cell Proteomics 7(4), 645–660 (2008).

12. Ishitamara, Y. et al. Enhancement of the efficiency of phosphoproteomic identification by removing phosphates after phosphopeptide enrichment. J. Proteome Res. 6(3), 1139–1144 (2007).

13. Saadaoui, B. et al. Combining proteomic tools to characterize the protein fraction of llama (Lama glama) milk. Electrophoresis 35, 1406–1418 (2014).

14. Ferranti, P. et al. Combining different proteomic approaches to resolve complexity of the milk protein fraction of dromedary, Bactrian camels and hybrids, from different regions of Kazakhstan. PLoS ONE 13(5), e0197026 (2018).

15. Henry, C., Saadaoui, B., Bouvier, F. & Cebo, C. Phosphoproteomics of the goat milk fat globule membrane: New insights into lipid droplet secretion from the mammary epithelial cell. Proteomics 15(13), 2307–2317 (2015).

16. Olumee-Shabon, Z. & Boehmer, J. L. Detection of casein phosphopeptides in goat milk. J. Proteome Res. 12(6), 3034–3041 (2013).

17. Ciavardelli, D., Sacchetta, P., Federici, G., Di Ilio, C. & Urbani, A. Protein phosphorylation stoichiometry by simultaneous ICP-QMS determination of phosphorus and sulfur oxide ions: a multivariable optimization of plasma operating conditions. Talanta 80, 1513–25 (2010).

18. Matéos, A. L., Miclo, D., Mollé, A., Dary, J. M. & Girardet, J. L. Equine αS1-casein: characterization of alternative splicing isoforms and determination of phosphorylation levels. J. Dairy Sci. 92, 3640–3615 (2009).

19. Matéos, A. L., Miclo, D., Mollé, A., Dary, J. M. & Girardet, J. L. Structure and function of milk protein genes. J. Protein Chem. 188, 329–352 (1998).

20. Olumee-Shabon, Z. & Boehmer, J. L. Detection of casein phosphopeptides in goat milk. J. Proteome Res. 12(6), 3034–3041 (2013).

21. Ciavardelli, D., Sacchetta, P., Federici, G., Di Ilio, C. & Urbani, A. Protein phosphorylation stoichiometry by simultaneous ICP-QMS determination of phosphorus and sulfur oxide ions: a multivariable optimization of plasma operating conditions. Talanta 80, 1513–25 (2010).

22. Matéos, A. L., Miclo, D., Mollé, A., Dary, J. M. & Girardet, J. L. Equine αS1-casein: characterization of alternative splicing isoforms and determination of phosphorylation levels. J. Dairy Sci. 92, 3640–3615 (2009).

23. Ferranti, P. et al. The primary structure of water buffalo α-s and β-casein: identification of phosphorylation sites and characterization of a novel β-casein variant. J. Protein Chem. 17, 835–44 (1998).

24. Park, Y. S. & Nam, M. S. Bioactive peptides in milk and dairy products: a review. Korean J. Food Sci. An. 35(6), 831–840 (2015).

25. Park, Y., Juárez, M., Ramos, M. & Haenlein, G. Physico-chemical characteristics of goats and sheep milk. Small Ruminant Res. 68, 88–113 (2007).

26. D’Aluria, E. et al. Precision medicine in cow’s milk allergy: proteomics perspectives from allergens to patients. J. Proteomics 100, 173–180 (2018).

27. O’Donnell, R., Holland, J. W., Deeth, H. C. & Alewpood, P. Milk proteomics. Int. Dairy J. 14, 1013–1023 (2004).

28. Khoury, G. A., Baliban, R. C. & Floudas, C. A. Proteome-wide post-translational modification statistics: frequency analysis and curation of the Swiss-Prot database. Sci. Rep. 1, 90 (2011).

29. Warden, S. M. et al. Post-translational modifications of the beta-1 subunit of AMP-activated protein kinase affect enzyme activity and cellular localization. Biochem. J. 354, 275–283 (2001).

30. Marvin, L. F., Parisod, V., Fay, L. B. & Guy, P. A. Characterization of lactosylated proteins of infant formula powders using two-dimensional gel electrophoresis and nanoelectrospray mass spectrometry. Proteomics 13, 2050–2512 (2002).

31. Charlwood, J. et al. Use of proteomic methodology for the characterization of human milk fat globular membrane proteins. Anal. Biochem. 301, 314–324 (2002).

32. Baeker, R., Haeberl, S., Schlatterer, K. & Schlatterer, B. Lipocalin-type prostaglandin D synthase in milk: A new biomarker for bovine mastitis. Prostaglandins Other Lipid Mediat. 67, 75–88 (2002).

33. West, D. Structure and function of phosphorylated residues of casein. J. Dairy Res. 53, 333–352 (1986).

34. Mercier, J. C. & Villette, J. L. Structure and function of milk protein genes. J. Dairy Sci. 76(10), 3079–3098 (1993).

35. Galvani, M., Hamdan, M. & Righetti, P. G. Two-dimensional gel electrophoresis/matrix assisted laser desorption/ionisation mass spectrometry of a milk powders. Rapid Commun. Mass Spectrom 14, 1889–1897 (2000).

36. Roncada, P. et al. Identification of caseins in goat milk. Proteomics 2(6), 723–726 (2002).

37. Fang, Z. H. et al. Genetic and nongenetic factors contributing to differences in αs-casein phosphorylation isoforms and other major milk proteins. J. Dairy Sci. 100, 5564–5577 (2017).

38. Murakami, K., Lagarde, M. & Yuki, Y. Identification of minor proteins of human colostrum and mature milk by two-dimensional electrophoresis. Electrophoresis 19, 2521–2527 (1998).

39. Yamada, M., Murakami, K., Wallengren, J. C. & Yuki, Y. Identification of low-abundance proteins of bovine colostral and mature milk using two-dimensional electrophoresis followed by microsequencing and mass spectrometry. Electrophoresis 23, 1153–1160 (2002).

40. Bijl, E., van Valenberg, H. J. E., Hupperta, T., van Hoofdonk, A. C. M. & Bovenhuis, H. Phosphorylation of αS1-casein is regulated by different genes. J. Dairy Sci. 97, 7240–7246 (2014).

41. Bijl, E. et al. Chymosin induced hydrolysis of caseins: Influence of degree of phosphorylation of alpha-S1-casein and genetic variants of beta-casein. Int. Dairy J. 39, 215–221 (2014).
42. Mamone, G. et al. Casein phosphoproteome: identification of phosphoproteins by combined mass spectrometry and two-dimensional gel electrophoresis. Electrophoresis 24, 2824–2837 (2003).
43. Chiannese, L. et al. Proteomic characterization of donkey milk ‘caseome’. J. Chromatogr. A 1217, 4834–4840 (2010).
44. Verma, M., Dige, M. S., Gautam, D., De, S. & Rout, P. K. Functional milk proteome analysis of genetically diverse goats from different agro climatic regions. J. Proteomics 227, 10396 (2020).
45. Olumee-Shabon, Z., Swain, T., Smith, E. A., Tall, E. & Boehmer, J. L. Proteomic analysis of differentially expressed proteins in caprine milk during experimentally induced endotoxin mastitis. J. Dairy Sci. 96, 2903–2912 (2013).
46. Sun, Y., Wang, C., Sun, X. & Guo, M. Comparative proteomics of whey and milk fat globule membrane proteins of Guanzhong goat and Holstein cow mature milk. J. Food Sci. 84(2), 244–253 (2019).
47. Cunsolo, V. et al. Zeus, Aesculapius, Amalthea and the proteome of goat milk. J. Proteome 128, 69–82 (2015).
48. Cebo, C., Caillat, H., Bouvier, F. & Martin, P. Major proteins of the goat milk fat globule membrane. J. Dairy Sci. 93, 868–876 (2009).
49. Anagnostopoulou, A. K. et al. Milk of Greek sheep and goat breeds; characterization by means of proteomics. J. Proteome 147, 76–84 (2016).
50. Farrell, H. M. Jr. Casein nomenclature, structure, and association. In The Encyclopedia of Dairy Sciences 2nd edn (eds Fuquay, J. W. et al.) 765–771 (Academic Press, London, 2011).
51. Cuello, M. et al. Toward milk speication through the monitoring of casein proteotypic peptides. Rapid Commun. Mass Spectrom. 24, 1687–1696 (2010).
52. Guarino, C. et al. Pepidomic approach, based on liquid chromatography/electrospray ionization tandem mass spectrometry, for detecting sheep’s milk in goat’s and cows cheeses. Rapid Commun. Mass Sp. 24, 3567–3577 (2010).
53. Sforza, S. et al. Cheese pepidomics: A detailed study on the evolution of the oligopeptide fraction in Parmigiano Reggiano cheese from curb to 24 months of aging. J. Dairy Sci. 95, 3514–3526 (2012).
54. Thomas, A. Towards a functional understanding of protein n-terminal acetylation. PLoS Biol. 9(5), e1001074 (2011).
55. Koehler, P. et al. Improved identification of wheat gluten proteins through alkylation of cysteine residues and peptide-based mass spectrometry. Sci. Rep. 3, 2279 (2013).
56. Holt, C., Carver, J. A., Ecroyd, H. & Thorn, D. C. Invited review: Caseins and the casein micelle: their biological functions, structures, and behavior in foods. J. Dairy Sci. 96, 6127–6146 (2013).
57. Lemmon, M. A. & Schlessinger, J. Cell signaling by receptor tyrosine kinases. Cell 141(7), 1117–1134 (2010).
58. Michaelidou, A. M. Factors influencing nutritional and health profile of milk and milk products. Small Rumin Res. 79(1), 42–50 (2008).
59. Karami, Z. & Akbari-adergani, B. Bioactive food derived peptides: a review on correlation between structure of bioactive peptides and their functional properties. J. Food Sci. Technol. 56(2), 535–547 (2019).
60. Grosclaude, F., Mahe, M. F., Brignon, G., Di-Stasio, L. & Jeunet, R. A Mendelian polymorphism underlying quantitative variations of goat aS1-casein. Genet. Sel. Evol 19, 399–412 (1987).
61. Grosclaude, F. et al. gene au fromage: le polymorphismie de la caseine α-S1 caprine, ses effets, son evolution. INRA Prod. Animales 7, 3–19 (1994).
62. Vassal, L., Delacroix-Buchet, A. & Bouillon, J. Influence des variants AA, EE, et FF de la caseine αs1caprine sur le rendement fromager et les caracteristiques sensorielles de fromages traditionnels: premiere observations. Lait 74, 89–103 (1994).
63. Sacchi, P. et al. Casein haplotype structure in five Italian goat breeds. J. Dairy Sci 88(4), 1561–1568 (2005).
64. Bevilacqua, C. et al. Interallelic recombination is probably responsible for the occurrence of a new a1-casein variant found in the goat species. Eur. J. Biochem. 269, 1293–1303 (2002).
65. Kumar, A., Rout, P. K., Mandal, A. & Roy, R. Identification of the CSN1S1 allele by PCR-RFLP method in Indian Goats. Animal 1, 1099–1104 (2007).
66. Verma, M. et al. Milk composition traits in Janunapari goats: genetic parameter estimation and effect of allelic variation in CSN1S1 gene. Int. J. Dairy Technol. 73(1), 12–21 (2020).
67. Manfredi, E., Ricordeau, G., Barbieri, M. E., Amigues, Y. & Bibe, B. Genotype at the a1-casein locus and selection of bucks on progeny test in the alpine and Saanen breeds. Genet. Sel. Evol. 27, 451–458 (1995).
68. Grosclaude, F. & Martin, P. Casein polymorphism in the goat. In: Proceeding of the IDF Seminar Held in Palmerston North, New Zealand session IV, pp. 241–253 (1997).
69. Nielsen, D. S., 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. 232, 673–682 (2017).

Acknowledgements
This research is supported by National Agriculture Science Fund (NASF), India Council of Agricultural Research (ICAR), ICAR- CIRG, India. Authors are thankful to Director, CIRG for providing necessary facilities to carry out the work.

Author contributions
P. K. R.: Concept development, experimental design, laboratory analysis: Sample collection, milk protein variant analysis, 2DE, nLC-MS/MS, data analysis and manuscript writing- original draft, reviewing and editing. M. V.: Laboratory experiments: Sample collection, milk protein variant analysis, 2DE, nLC-MS/MS, software data analysis and manuscript writing-original draft, reviewing and editing.

Competing interest
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-85094-9.

Contact and requests for materials should be addressed to P.K.R.

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.
