DATASET BRIEF

# A Bovine PeptideAtlas of milk and mammary gland proteomes

*Stine L. Bislev[1], Eric W. Deutsch[2], Zhi Sun[2], Terry Farrah[2], Ruedi Aebersold[3,4,5], Robert L. Moritz[2], Emøke Bendixen[1] and Marius C. Codrea[1]*

[1] Faculty of Science and Technology, Department of Animal Science and Department of Molecular Biology and Genetics, Aarhus University, Aarhus, Denmark
[2] Institute for Systems Biology, Seattle, WA, USA
[3] Institute of Molecular Systems Biology, ETH Zurich, Zurich, Switzerland
[4] Faculty of Science, University of Zurich, Zurich, Switzerland
[5] Competence Center for Systems Physiology and Metabolic Disease, ETH Zurich, Zurich, Switzerland

Received: February 14, 2012
Revised: June 4, 2012
Accepted: July 5, 2012

Proteome information resources of farm animals are lagging behind those of the classical model organisms despite their important biological and economic relevance. Here, we present a Bovine PeptideAtlas, representing a first collection of *Bos taurus* proteome data sets within the PeptideAtlas framework. This database was built primarily as a source of information for designing selected reaction monitoring assays for studying milk production and mammary gland health, but it has an intrinsic general value for the farm animal research community. The Bovine PeptideAtlas comprises 1921 proteins at 1.2% false discovery rate (FDR) and 8559 distinct peptides at 0.29% FDR identified in 107 samples from six tissues. The PeptideAtlas web interface has a rich set of visualization and data exploration tools, enabling users to interactively mine information about individual proteins and peptides, their prototypic features, genome mappings, and supporting spectral evidence.

**Keywords:**
Animal proteomics / *Bos taurus* / Mammary gland proteome / Milk proteome / PeptideAtlas / Proteotypic peptides

Publicly available data repositories such as PRIDE [1], the Global Proteome Machine Database (GPMDB) [2], Tranche [3], and PeptideAtlas [4] play a fundamental role in successful development of new methods and progress in biological sciences. A recent review [5] presents the roles of several well-established MS proteomics repositories and highlights the need for further diversity of data. However, in these repositories, organisms that are not commonly regarded as classical model organisms are widely underrepresented, if represented at all. Studies of farm animal proteomes have recently gained much interest, particularly regarding cattle and pig. This is mainly because proteomics offers unmatched opportunities to characterize biological traits of farm animals, and is thereby a key to improve industrial production of meat and milk. This is of immediate relevance for economic gain as well as for alleviating problems related to animal welfare and product quality within agriculture and food industries [6]. While individual studies provide valuable new insights into theproteomic mechanisms (e.g. of particular diseases), farm animal data repositories are by far lagging behind. For example, as of November 2011, the PRIDE database (http://www.ebi.ac.uk/pride) [1] contained more than 3320 human proteome data sets, but only 34 bovine experiments, mainly originating from studies of reproduction biology characterizing sperm and oocyte proteomes.

Within the past decade, the PeptideAtlas project (http://www.peptideatlas.org) has provided a large-scale

**Correspondence:** Dr. Emøke Bendixen, Faculty of Science and Technology, Department of Molecular Biology and Genetics, Aarhus University, Gustav Wieds Vej 10C, 8000 Aarhus C, Denmark
**E-mail:** emoke.bendixen@agrsci.dk
**Fax:** +45-8612-3178

**Abbreviations: ESS**, Empirical Suitability Score; **FDR**, false discovery rate; **GPMDB**, Global Proteome Machine Database; **PSM**, peptide-spectrum matches; **PSS**, Predicted Suitability Score; **SRM**, Selected Reaction Monitoring

**Table 1.** An overview of the tissue coverage

| Tissues | Number of experiments | Number of canonical proteins at 1% FDR | Number of peptides at 0.2% PSM FDR |
|---|---|---|---|
| Mammary epithelial cells | 21 | 1061 | 3473 |
| Colostrum | 12 | 342 | 1003 |
| Milk | 24 | 747 | 2061 |
| Udder | 25 | 803 | 2778 |
| Hoof | 11 | 385 | 1042 |
| Mitochondria | 14 | 1081 | 3656 |
| Total (unique) | 107 | 1921 | 8559 |

assembly of LC-MS/MS-based shotgun proteome data, and covers the proteomes of many species, most importantly those of human, including specialized builds for specific tissues and body fluids, and common model organisms such as yeast (*Saccharomyces cerevisiae*) and *Drosophila melanogaster* [7–9]. The PeptideAtlas has been useful in characterizing the biological systems of these organisms as well as in setting foundations for other species such as the honeybee [10]. The PeptideAtlas has become the tool of choice for selecting proteotypic peptides [5], which can be used to build methods for targeted proteomics and Selected Reaction Monitoring (SRM) [4]. The SRM method has recently become a widely used approach for detecting low-abundance proteins in cells and body fluids, and addresses the problems of analyzing a large variety of proteins present at both high and low abundance within complex biological samples [11].

Milk is a complex body fluid that includes a wide range of secreted proteins, hence may provide diagnostic measures and reflects the health state of animals. Moreover, bovine milk is collected daily, providing easy access to routine diagnostics. But the very large dynamic range (caseins account for 80% of the total protein content in milk) complicates analyses of specific proteins; hence, developing targeted methods for routine analyses of specific milk proteins of potential diagnostic value provides a promising approach.

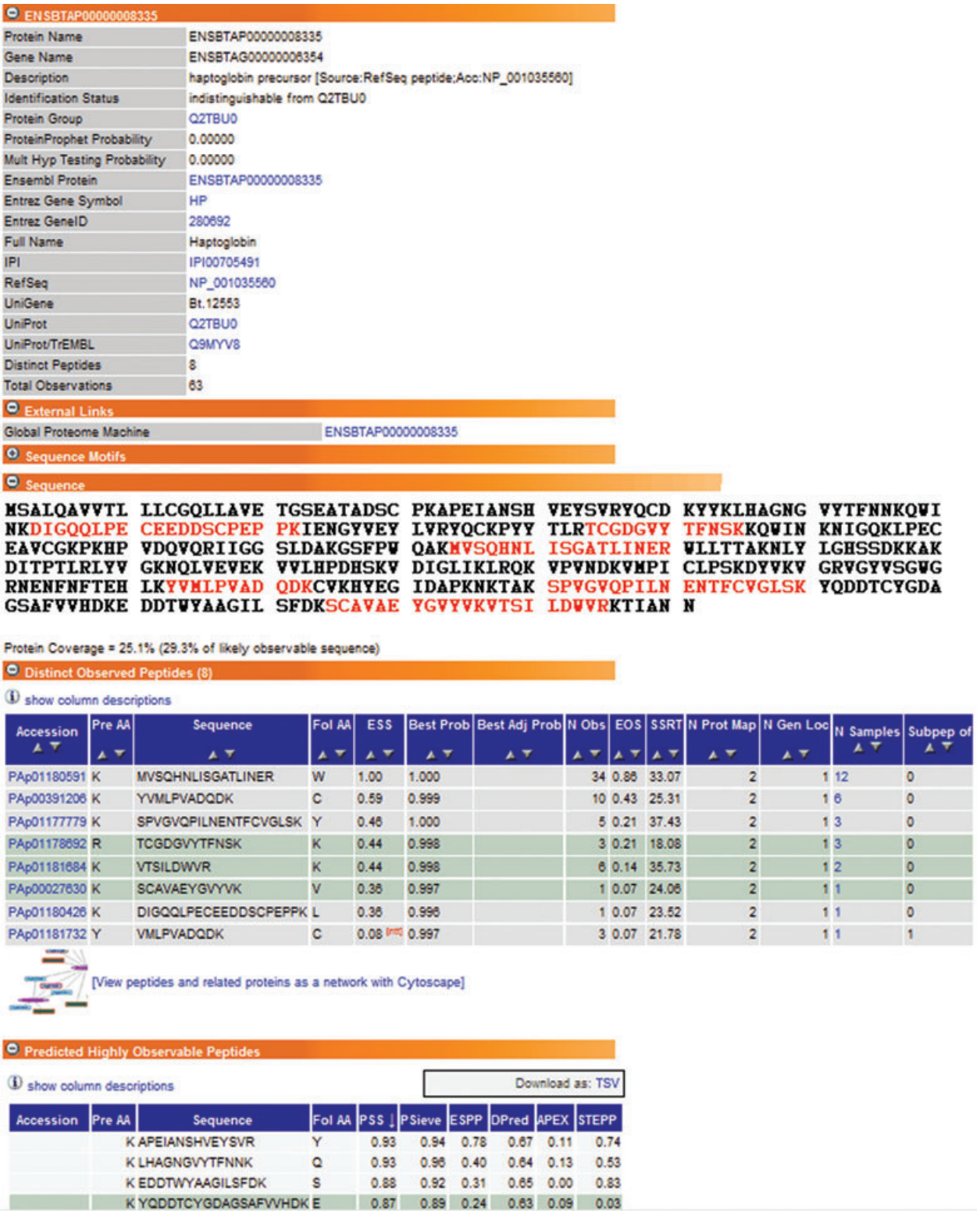
Here, we present a Bovine PeptideAtlas that covers tissues and body fluids relevant for milk and mammary gland proteomes (Build: Cow Milk 2011–12, https://db.systemsbiology.net/sbeams/cgi/PeptideAtlas/buildDeta-ils?atlas_build_id=320). For the purpose of building the PeptideAtlas, we collected a set of representative samples from various in-house cattle proteome projects, aiming to provide good protein and peptide coverage on inflammatory and host response proteins. In addition to milk samples, we included colostrum, the bovine mammary epithelial cell line MAC-T used for in vitro studies of mammary epithelial host response, udder tissues, a subcellular mitochondria fraction, and hoof tissues (see Table 1). We included hoof tissues because these are major sites of inflammation in all hoofed animals, hence expected to provide information of proteins related to inflammatory control [12]. Further details of the samples can be found on http://www.peptideatlas.org/repository when selecting cow as organism.

All samples were collected, prepared, and analyzed by shotgun 2D LC-MS/MS, according to our protocol described in full details in our previous papers, e.g. milk samples [13] and mammary gland tissue samples [14]. In summary, proteins were extracted and protein concentrations were determined. Cysteine residues were reduced and blocked and proteins were digested with trypsin (1:10 w/w). In some of the samples, peptides were labeled with the iTRAQ™ Reagent MultiPlex Kit according to manufacturer's manual (ABSciex, Foster City, CA, USA). Tryptic peptides were separated by strong cation exchange chromatography, followed by RP chromatography and analyzed on the quadrupole-TOF mass spectrometer Q-star Elite (ABSciex).

The construction of the Bovine PeptideAtlas followed the pipeline described in [15]. In short, the raw data files were converted from the binary wiff format to mzML format [16] with the msconvert tool from ProteoWizard [17], which used Protein Pilot 3.0 (ABSciex) libraries for peak detection and charge state determination. The mzML files were searched with the X!Tandem [18] sequence search engine with the *k*-score plug-in [19]. The sequence database used for searching was compiled as a nonredundant union of bovine sequences from UniProt, Ensembl, and UniGene plus the cRAP contaminants (http://www.thegpm.org/crap/index.html). A like number of decoy sequences were appended to the target protein sequences. The results of each search were processed through the Trans-Proteomic Pipeline [20] to yield a list of high-confidence peptide identifications. ProteinProphet [21] was run on all data sets combined to generate protein identifications and protein groupings, which were then refined and classified according to the Cedar scheme [22]. All peptide sequences were mapped to the Ensembl (http://www.ensembl.org) version 56 build, thereby allowing us to calculate chromosomal coordinates for peptides found in Ensembl; for the peptides that do not map to Ensembl sequences, chromosomal coordinates are not available. The raw data, search parameters, and search database are downloadable at http://www.peptideatlas.org/repository.

The overall summary of the Bovine PeptideAtlas protein and peptide coverage of individual tissues is given in Table 1. The current release of the Bovine PeptideAtlas gives experimental peptide information for 8559 unique peptides at 0.29% peptide false discovery rate (FDR) and 0.2% peptide-spectrum matches (PSM) FDR. Under the Cedar protein identification scheme, these represent 1921 *canonical* (highly distinguishable and nonredundant) proteins at 1% FDR. Compared to most schemes, the Cedar canonical protein list provides a very conservative and trustworthy estimate of the number of distinct protein molecules observed. This number currently represents approximately 9% of the 22 000 predicted bovine proteins, based on the latest prediction of coding sequences from the completed bovine genome assembly [23]. A wide selection of pathways and groups of proteins well



ENSBTAP00000008335

| | |
|---|---|
| Protein Name | ENSBTAP00000008335 |
| Gene Name | ENSBTAG00000006354 |
| Description | haptoglobin precursor [Source:RefSeq peptide;Acc:NP_001035560] |
| Identification Status | indistinguishable from Q2TBU0 |
| Protein Group | Q2TBU0 |
| ProteinProphet Probability | 0.00000 |
| Mult Hyp Testing Probability | 0.00000 |
| Ensembl Protein | ENSBTAP00000008335 |
| Entrez Gene Symbol | HP |
| Entrez GeneID | 280692 |
| Full Name | Haptoglobin |
| IPI | IPI00705491 |
| RefSeq | NP_001035560 |
| UniGene | Bt.12553 |
| UniProt | Q2TBU0 |
| UniProt/TrEMBL | Q9MYV8 |
| Distinct Peptides | 8 |
| Total Observations | 63 |

External Links

| | |
|---|---|
| Global Proteome Machine | ENSBTAP00000008335 |

Sequence Motifs

Sequence

MSALQAVVTL LLCGQLLAVE TGSEATADSC PKAPEIANSH VEYSVRYQCD KYYKLHAGNG VYTFNNKQVI
NKDIGQQLPE CEEDDSCPEP PKIENGYVEY LVRYQCKPYY TLRTCGDGVY TFNSKKQVIN KNIGQKLPEC
EAVCGKPKHP VDQVQRIIGG SLDAKGSFPW QAKMVSQHNL ISGATLINER WLLTTAKNLY LGHSSDKKAK
DITPTLRLYV GKNQLVEVEK VVLHPDHSKV DIGLIKLRQK VPVNDKVMPI CLPSKDYVKV GRVGYVSGWG
RNENFNFTEH LKYVMLPVAD QDKCVKHYEG IDAPKNKTAK SPVGVQPILN ENTFCVGLSK YQDDTCYGDA
GSAFVVHDKE DDTWYAAGIL SFDKSCAVAE YGVYVKVTSI LDWVRKTIAN N

Protein Coverage = 25.1% (29.3% of likely observable sequence)

Distinct Observed Peptides (8)

show column descriptions

| Accession | Pre AA | Sequence | Fol AA | ESS | Best Prob | Best Adj Prob | N Obs | EOS | SSRT | N Prot Map | N Gen Loc | N Samples | Subpep of |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PAp01180591 | K | MVSQHNLISGATLINER | W | 1.00 | 1.000 | | 34 | 0.86 | 33.07 | 2 | 1 | 12 | 0 |
| PAp00391206 | K | YVMLPVADQDK | C | 0.59 | 0.999 | | 10 | 0.43 | 25.31 | 2 | 1 | 6 | 0 |
| PAp01177779 | K | SPVGVQPILNENTFCVGLSK | Y | 0.46 | 1.000 | | 5 | 0.21 | 37.43 | 2 | 1 | 3 | 0 |
| PAp01178692 | R | TCGDGVYTFNSK | K | 0.44 | 0.998 | | 3 | 0.21 | 18.08 | 2 | 1 | 3 | 0 |
| PAp01181684 | K | VTSILDWVR | K | 0.44 | 0.998 | | 6 | 0.14 | 35.73 | 2 | 1 | 2 | 0 |
| PAp00027630 | K | SCAVAEYGVYVK | V | 0.36 | 0.997 | | 1 | 0.07 | 24.06 | 2 | 1 | 1 | 0 |
| PAp01180426 | K | DIGQQLPECEEDDSCPEPPK | L | 0.36 | 0.996 | | 1 | 0.07 | 23.52 | 2 | 1 | 1 | 0 |
| PAp01181732 | Y | VMLPVADQDK | C | 0.08 | 0.997 | | 3 | 0.07 | 21.78 | 2 | 1 | 1 | 1 |

[View peptides and related proteins as a network with Cytoscape]

Predicted Highly Observable Peptides

show column descriptions

Download as: TSV

| Accession | Pre AA | Sequence | Fol AA | PSS | PSieve | ESPP | DPred | APEX | STEPP |
|---|---|---|---|---|---|---|---|---|---|
| | K | APEIANSHVEYSVR | Y | 0.93 | 0.94 | 0.78 | 0.67 | 0.11 | 0.74 |
| | K | LHAGNGVYTFNNK | Q | 0.93 | 0.96 | 0.40 | 0.64 | 0.13 | 0.53 |
| | K | EDDTWYAAGILSFDK | S | 0.88 | 0.92 | 0.31 | 0.65 | 0.00 | 0.83 |
| | K | YQDDTCYGDAGSAFVVHDK | E | 0.87 | 0.89 | 0.24 | 0.63 | 0.09 | 0.03 |

**Figure 1.** Example of a protein view in the Bovine PeptideAtlas for the protein haptoglobin (Hp). There are several collapsible sections, each of which provides a related set of information about the protein. The first section provides an overview of information about the protein in general. Next, observed peptides are shown in red on the full protein amino acid sequence, and more information about each peptide is provided in the section Distinct Observed Peptides. The section Predicted Highly Observable Peptides lists theoretical peptides for the protein digested in silico by different prediction software tools.

known to play active roles in inflammation and host defense are identified in this Bovine PeptideAtlas, including all commonly known acute-phase proteins, at least six members of the cathepsin superfamily, seven members of the antibacterial cathelicidin family, and more than 12 different lymphocyte surface antigens (CD antigens).

The PeptideAtlas interface allows the user to explore individual proteins. The Protein View page provides a fast overview of protein sequence coverage, observed peptides, and predicted observable peptides (Fig. 1). Both observed peptides and predicted observable peptides are ranked by suitability scores; Empirical Suitability Score (ESS) and Predicted Suitability Score (PSS), respectively. The ESS represents how suitable the peptide is as a proteotypic peptide and might help in the development of targeted proteomics experiments. For proteins not observed in this data set, for example low-abundance proteins, the PSS indicates which peptides might be easily detectable in an electrospray mass spectrometer. A detailed review of all the features of PeptideAtlas is given in [22]. The collection of data presented here is meant to provide a resource for future bovine proteomics projects, and in particular for selecting target peptides for use in SRM-based workflows. The Bovine PeptideAtlas provides for the first time such experimental evidence. Vizcaino et al. [5] emphasize the current joint efforts of the key players: PRIDE [1], PeptideAtlas [4], the GPMDB [2], and Tranche [3], in the ProteomeXchange consortium (http://www.proteomexchange.org). The release of the Bovine PeptideAtlas represents a substantial contribution for large-mammal model organisms.

*This work was supported by the BIOSENS Consortium project, the Danish Ministry of Food Agriculture and Fisheries, Lattec I/S, the Milk Levy Fund, the Faculty of Science and Technology, the Danish Strategic Research Council, and the Graduate School of Agriculture, Food and Environment at Aarhus University. This work was supported in part by American Recovery and Reinvestment Act (ARRA) funds through grant number RC2 HG005805 (to RM) from the National Human Genome Research Institute, National Institutes of Health, the Luxembourg Centre for Systems Biomedicine and the University of Luxembourg (to RM), the National Institute of General Medical Sciences, (grant No. GM087221 to EWD), and the National Heart, Lung, and Blood Institute, (contract No. N01-HV-28179 to RA).*

*The authors have declared no conflict of interest.*

## References

[1] Vizcaino, J. A., Cote, R., Reisinger, F., Foster, J. M. et al., A guide to the Proteomics Identifications Database proteomics data repository. *Proteomics* 2009, *9*, 4276–4283.

[2] Craig, R., Cortens, J. P., Beavis, R. C., Open source system for analyzing, validating, and storing protein identification data. *J. Proteome. Res.* 2004, *3*, 1234–1242.

[3] Falkner, J. A., Andrews, P. C., P6-T Tranche: secure decentralized data storage for the proteomics community. *J. Biomol. Tech* 2007, *18*, 3.

[4] Deutsch, E. W., Lam, H., Aebersold, R., PeptideAtlas: a resource for target selection for emerging targeted proteomics workflows. *EMBO Rep.* 2008, *9*, 429–434.

[5] Vizcaino, J. A., Foster, J. M., Martens, L., Proteomics data repositories: providing a safe haven for your data and acting as a springboard for further research. *J. Proteomics.* 2010, *73*, 2136–2146.

[6] Bendixen, E., Danielsen, M., Hollung, K., Gianazza, E. et al., Farm animal proteomics—a review. *J. Proteomics* 2011, *74*, 282–293.

[7] Desiere, F., Deutsch, E. W., King, N. L., Nesvizhskii, A. I. et al., The PeptideAtlas project. *Nucleic Acids Res.* 2006, *34*, D655–D658.

[8] King, N. L., Deutsch, E. W., Ranish, J. A., Nesvizhskii, A. I. et al., Analysis of the *Saccharomyces cerevisiae* proteome with PeptideAtlas. *Genome Biol.* 2006, *7*, R106.

[9] Loevenich, S. N., Brunner, E., King, N. L., Deutsch, E. W. et al., The *Drosophila melanogaster* PeptideAtlas facilitates the use of peptide data for improved fly proteomics and genome annotation. *BMC. Bioinformatics* 2009, *10*, 59.

[10] Chan, Q. W., Parker, R., Sun, Z., Deutsch, E. W. et al., A honey bee (*Apis mellifera* L.) PeptideAtlas crossing castes and tissues. *BMC. Genomics* 2011, *12*, 290.

[11] Picotti, P., Bodenmiller, B., Mueller, L. N., Domon, B. et al., Full dynamic range proteome analysis of *S. cerevisiae* by targeted proteomics. *Cell* 2009, *138*, 795–806.

[12] Danscher, A. M., Toelboell, T. H., Wattle, O. Biomechanics and histology of bovine claw suspensory tissue in early acute laminitis. *J. Dairy Sci.* 2010, *93*, 53–62.

[13] Danielsen, M., Codrea, M. C., Ingvartsen, K. L., Friggens, N. C. et al., Quantitative milk proteomics—host responses to lipopolysaccharide-mediated inflammation of bovine mammary gland. *Proteomics* 2010, *10*, 2240–2249.

[14] Bislev, S. L., Kusebauch, U., Codrea, M. C., Beynon, R. J. et al., Quantotypic properties of QconCAT peptides targeting bovine host response to *Streptococcus uberis*. *J. Proteome Res.* 2012, *11*, 1832–1843.

[15] Farrah, T., Deutsch, E. W., Omenn, G. S., Campbell, D. S. et al., A high-confidence human plasma proteome reference set with estimated concentrations in PeptideAtlas. *Mol. Cell Proteomics* 2011, *10*, M110.

[16] Martens, L., Chambers, M., Sturm, M., Kessner, D. et al., mzML—a community standard for mass spectrometry data. *Mol. Cell Proteomics* 2011, *10*, R110.

[17] Kessner, D., Chambers, M., Burke, R., Agus, D. et al., ProteoWizard: open source software for rapid proteomics tools development. *Bioinformatics* 2008, *24*, 2534–2536.

[18] Craig, R., Beavis, R. C., TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* 2004, *20*, 1466–1467.

[19] MacLean, B., Eng, J. K., Beavis, R. C., McIntosh, M., General framework for developing and evaluating database scoring algorithms using the TANDEM search engine. *Bioinformatics* 2006, *22*, 2830–2832.

[20] Keller, A., Eng, J., Zhang, N., Li, X. J. et al., A uniform proteomics MS/MS analysis platform utilizing open XML file formats. *Mol. Syst. Biol.* 2005, *1*, 1–8.

[21] Nesvizhskii, A. I., Keller, A., Kolker, E., Aebersold, R., A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* 2003, *75*, 4646–4658.

[22] Farrah, T., Deutsch, E. W., Aebersold, R., Using the human plasma PeptideAtlas to study human plasma proteins. *Methods Mol. Biol.* 2011, *728*, 349–374.

[23] Elsik, C. G., Tellam, R. L., Worley, K. C., Gibbs, R. A. et al., The genome sequence of taurine cattle: a window to ruminant biology and evolution. *Science* 2009, *324*, 522–528.