Published in final edited form as:

Proteomics. 2012 September ; 12(18): 2895–2899. doi:10.1002/pmic.201200057.

A Bovine PeptideAtlas of milk and mammary gland proteomes

Stine L. Bislev¹, Eric W. Deutsch², Zhi Sun², Terry Farrah², Ruedi Aebersold^{3,4,5}, Robert L. Moritz², Emøke Bendixen¹, and Marius C. Codrea^{1,*}

¹ Department of Animal Science, Faculty of Science and Technology, Aarhus University, 8830 Tjele, Denmark ² Institute for Systems Biology, Seattle, Washington 98109, USA ³ Institute of Molecular Systems Biology, ETH Zurich, 8093 Zurich, Switzerland ⁴ Faculty of Science, University of Zurich, 8057 Zurich, Switzerland ⁵ Competence Center for Systems Physiology and Metabolic Disease, ETH Zurich, 8093 Zurich, Switzerland

Abstract

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 bovine proteomics datasets 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 6 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 genome mappings, and supporting spectral evidence.

Keywords

Bos Taurus; mammary gland proteome; milk proteome; PeptideAtlas; proteotypic peptides

Publicly available data repositories like PRIDE [1], the Global Proteome Machine Database [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

* Correspondence: Dr. Emøke Bendixen, Department of Animal Science, Faculty of Science and Technology, Aarhus University, Blichers Allé, P.O. Box 50, DK-8830 Tjele, Denmark, emoke.bendixen@agrsci.dk, Phone number: + 45 8715 8001, Fax: + 45 8715 6000.

The authors have declared no conflict of interest.

problems related to animal welfare and product quality within agriculture and food industries [6]. While individual studies provide valuable new insights into the proteomic mechanisms (e.g., of particular diseases), farm animal data repositories are by far lagging behind. For example, as of November 2011 the PRIDE database (www.ebi.ac.uk/pride) [1] contained more than 3320 human proteome datasets, but only 34 bovine experiments, mainly originating from studies of reproduction biology characterizing sperm and oocyte proteomes.

Within the past decade, the PeptideAtlas project (www.peptideatlas.org) has provided a large scale 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 like 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 harbors/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/buildDetails?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 they are a major site for 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 in a similar way, according to our protocol described in full details in several of 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 Multi-Plex Kit according to manufacturer's manual (ABSciex, Foster City, CA, USA). Tryptic peptides were separated by strong cation exchange chromatography, followed by reverse-phase chromatography and analyzed on the Q-TOF mass spectrometer Q-star Elite (ABSciex, Foster City, CA, USA).

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, Foster City, CA, USA) 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 non-redundant union of bovine sequences from UniProt, Ensembl and UniGene plus the cRAP contaminants (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 datasets 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 (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 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 FDR and 0.2% PSM FDR. Under the Cedar protein identification scheme, these represent 1921 *canonical* (highly distinguishable and non-redundant) 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 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 6 members of the cathepsin superfamily, 7 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 dataset, 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 Global Proteome Machine Database (GPMDB) [2] and Tranche [3], in the ProteomeXchange consortium (www.proteomexchange.org). The release of the Bovine PeptideAtlas represents a substantial contribution for large-mammal model organisms.

Acknowledgments

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. Funding has also been provided by the National Institutes of Health-National Human Genome Research Institute (grant No. HG005805 to R.M.), the Luxembourg Centre for Systems Biomedicine and the University of Luxembourg (to R.M.), the National Institute of General Medical Sciences, (grant No. GM087221 to E.W.D), and the National Heart, Lung, and Blood Institute, (contract No. N01-HV-28179 to R.A.).

Abbreviations

GPMDB	the Global Proteome Machine Database
ESS	Empirical Suitability Score
PSS	Predicted Suitability Score
FDR	False discovery rate
PSM	peptide-spectrum matches

References

1. Vizcaino JA, Cote R, Reisinger F, Foster JM, et al. A guide to the Proteomics Identifications Database proteomics data repository. *Proteomics*. 2009; 9:4276–4283. [PubMed: 19662629]
2. Craig R, Cortens JP, Beavis RC. Open source system for analyzing, validating, and storing protein identification data. *J. Proteome. Res.* 2004; 3:1234–1242. [PubMed: 15595733]
3. Falkner JA, Andrews PC. P6-T Tranche: Secure Decentralized Data Storage for the Proteomics Community. *J. Biomol. Tech.* 2007; 18:3.
4. Deutsch EW, Lam H, Aebersold R. PeptideAtlas: a resource for target selection for emerging targeted proteomics workflows. *EMBO Rep.* 2008; 9:429–434. [PubMed: 18451766]
5. Vizcaino JA, Foster JM, 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. [PubMed: 20615486]
6. Bendixen E, Danielsen M, Hollung K, Gianazza E, et al. Farm animal proteomics--a review. *J. Proteomics*. 2011; 74:282–293. [PubMed: 21112346]
7. Desiere F, Deutsch EW, King NL, Nesvizhskii AI, et al. The PeptideAtlas project. *Nucleic Acids Res.* 2006; 34:D655–D658. [PubMed: 16381952]
8. King NL, Deutsch EW, Ranish JA, Nesvizhskii AI, et al. Analysis of the *Saccharomyces cerevisiae* proteome with PeptideAtlas. *Genome Biol.* 2006; 7:R106. [PubMed: 17101051]
9. Loevenich SN, Brunner E, King NL, Deutsch EW, et al. The *Drosophila melanogaster* PeptideAtlas facilitates the use of peptide data for improved fly proteomics and genome annotation. *BMC. Bioinformatics*. 2009; 10:59. [PubMed: 19210778]

10. Chan QW, Parker R, Sun Z, Deutsch EW, et al. A honey bee (*Apis mellifera* L.) PeptideAtlas crossing castes and tissues. *BMC. Genomics*. 2011; 12:290. [PubMed: 21639908]
11. Picotti P, Bodenmiller B, Mueller LN, Domon B, et al. Full dynamic range proteome analysis of *S. cerevisiae* by targeted proteomics. *Cell*. 2009; 138:795–806. [PubMed: 19664813]
12. Danscher AM, Toelboell TH, Wattle O. Biomechanics and histology of bovine claw suspensory tissue in early acute laminitis. *J. Dairy Sci*. 2010; 93:53–62. [PubMed: 20059904]
13. Danielsen M, Codrea MC, Ingvarsen KL, Friggens NC, et al. Quantitative milk proteomics--host responses to lipopolysaccharide-mediated inflammation of bovine mammary gland. *Proteomics*. 2010; 10:2240–2249. [PubMed: 20352626]
14. Bislev SL, Kusebauch U, Codrea MC, Beynon RJ, et al. Quantotypic Properties of QconCAT Peptides Targeting Bovine Host Response to *Streptococcus uberis*. *J. Proteome. Res*. 2012; 11:1832–1843. [PubMed: 22256911]
15. Farrah T, Deutsch EW, Omenn GS, Campbell DS, et al. A high-confidence human plasma proteome reference set with estimated concentrations in PeptideAtlas. *Mol. Cell Proteomics*. 2011; 10:M110. [PubMed: 21632744]
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. [PubMed: 20716697]
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. [PubMed: 18606607]
18. Craig R, Beavis RC. TANDEM: matching proteins with tandem mass spectra. *Bioinformatics*. 2004; 20:1466–1467. [PubMed: 14976030]
19. MacLean B, Eng JK, Beavis RC, McIntosh M. General framework for developing and evaluating database scoring algorithms using the TANDEM search engine. *Bioinformatics*. 2006; 22:2830–2832. [PubMed: 16877754]
20. Keller A, Eng J, Zhang N, Li XJ, et al. A uniform proteomics MS/MS analysis platform utilizing open XML file formats. *Mol. Syst. Biol*. 2005; 1:2005. [PubMed: 16729052]
21. Nesvizhskii AI, Keller A, Kolker E, Aebersold R. A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem*. 2003; 75:4646–4658. [PubMed: 14632076]
22. Farrah T, Deutsch EW, Aebersold R. Using the Human Plasma PeptideAtlas to study human plasma proteins. *Methods Mol. Biol*. 2011; 728:349–374. [PubMed: 21468960]
23. Elsik CG, Tellam RL, Worley KC, Gibbs RA, et al. The genome sequence of taurine cattle: a window to ruminant biology and evolution. *Science*. 2009; 324:522–528. [PubMed: 19390049]

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.

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