Data Article

Proteomic data of donkey's milk

Stavros C. Proikakis [a], Efterpi V. Bouroutzika [a,b],
Athanasios K. Anagnostopoulos [a], George Th. Tsangaris [a,*]

[a] *Proteomics Research Unit, Biomedical Research Foundation of the Academy of Athens, Athens 115 27, Greece*
[b] *Veterinary Faculty, University of Thessaly, Karditsa, Greece*

A R T I C L E   I N F O

A B S T R A C T

Donkey's milk has been recognized as milk of high biological value and it also has the closest composition to human milk. However, the total protein content of donkey's milk has not been adequately identified. The aim of this analysis is to investigate the proteomic content of that milk. Specific commercially available only milk was analyzed by "shotgun" proteomic methods to identify the proteins it contained in as much detail as possible. The application of the above approach resulted in the identification of a total of 633 different proteins, which were grouped based on their molecular function and their biological process. Furthermore, the proteins visualized graphically according to the GeneOntology (GO) system. The identified proteins confirm the high nutritional value of the donkey milk, governing future steps in optimizing its characteristic and uses.

---

* Corresponding author.
   E-mail address: gthtsangaris@bioacademy.gr (G.Th. Tsangaris).

Specifications Table

Subject	Omics: Proteomics
Specific subject area	Commercially available donkey milk
Type of data	Table, Figure
How data was acquired	1D-nanoLC-MS/MS, bottom-up proteomics
	Dionex Ultimate 3000 nanoHPLC system coupled to an LTQ Velos *Orbitrap Elite mass spectrometer* (Thermo Scientific, Rockford, IL, USA)
	PepMap® RSLC, C18, 100 Å, 3-μm-bead-packed 15-cm column and 2-μm-bead-packed 50-cm column (Thermo Scientific)
	Proteome Discoverer 1.4 software (Thermo Scientific), Sequest search engine searching the *Mammalia* (628 proteins), *Equus asinus* (5 proteins) *.fasta databases for milk of donkey.
Data format	Analyzed
Parameters for data collection	Commercially available pasteurized donkey milk samples from 3 different regions across Greece (Attika, Epirus, Macedonia) were collected and analyzed in order to characterize the complete protein content of the milk.
Description of data collection	6 donkey milk samples from 3 regions across Greece (2 from each region) had been purchased from super-markets and bio-shops.
Data source location	Institution: Biomedical Research Foundation Academy of Athens City/Town/Region: Athens Country: Greece
Data accessibility	With the article

Value of the Data

- A new report on the donkey milk whey proteome.
- The data can be used to manage future steps to optimize the characteristics of donkey milk products.
- The analysis of data may produce new dairy products.

1. Data Description

Donkey milk has been recognized as milk of high biological value because of its many antibacterial proteins [1,2]. In total 6 commercially available donkey milk samples were collected and analyzed in triplicate. Altogether 633 proteins were totally identified in the studied samples. In Supp. Table 1. identified proteins are shown by their accession number and their description according to

Swiss-Prot/UniprotKB database, while all the relevant information for their identification are included. In Supp. Table 2 are presented the 5 proteins which were identified in *Equus asinus* *.fasta database. In Fig. 1, identified proteins are classified based on their molecular function, in Fig. 2, they are classified based on their biological process and in Fig. 3 they classified according to their subcellular localization. All figures' data have been elaborated based on Swiss-Prot/UniprotKB database.

2. Experimental Design, Materials and Methods

2.1. Sample collection

Representative 6 samples in total of commercially available donkey's milk from 3 different regions across Greece (2 samples per region) were collected in deep frozen state. After the samples' thaw, aliquots of 2ml from each milk sample were taken and stored at −20 °C until analysis.
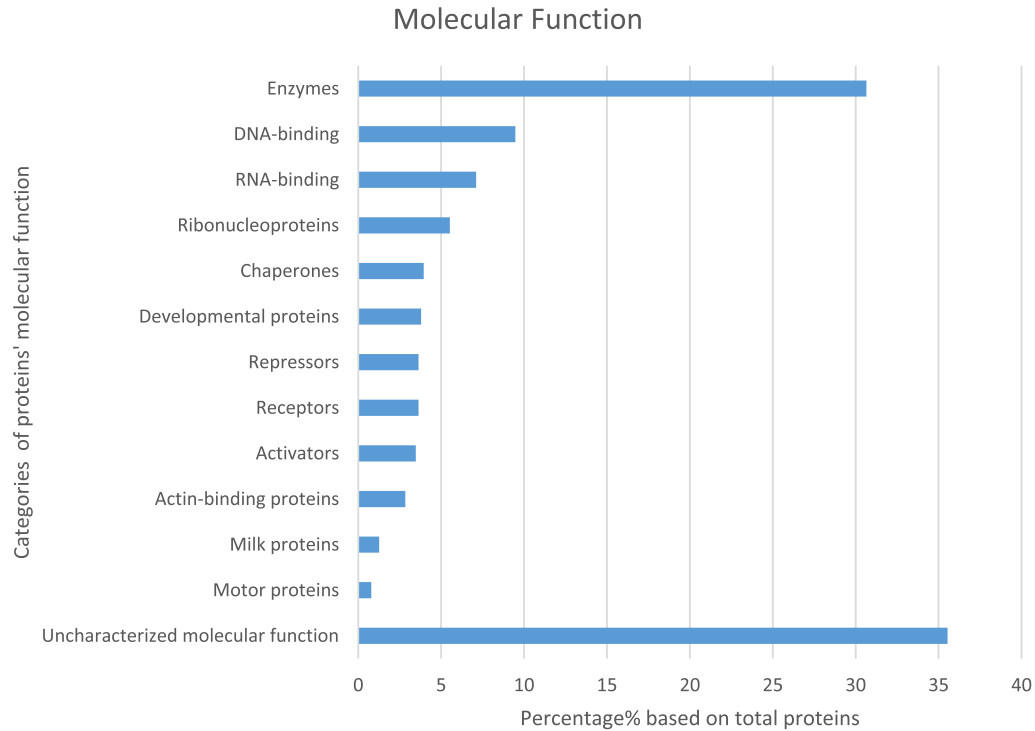
Fig. 1. Classification of donkey milk protein groups according to their molecular function. The molecular function of many proteins has not been characterized yet.
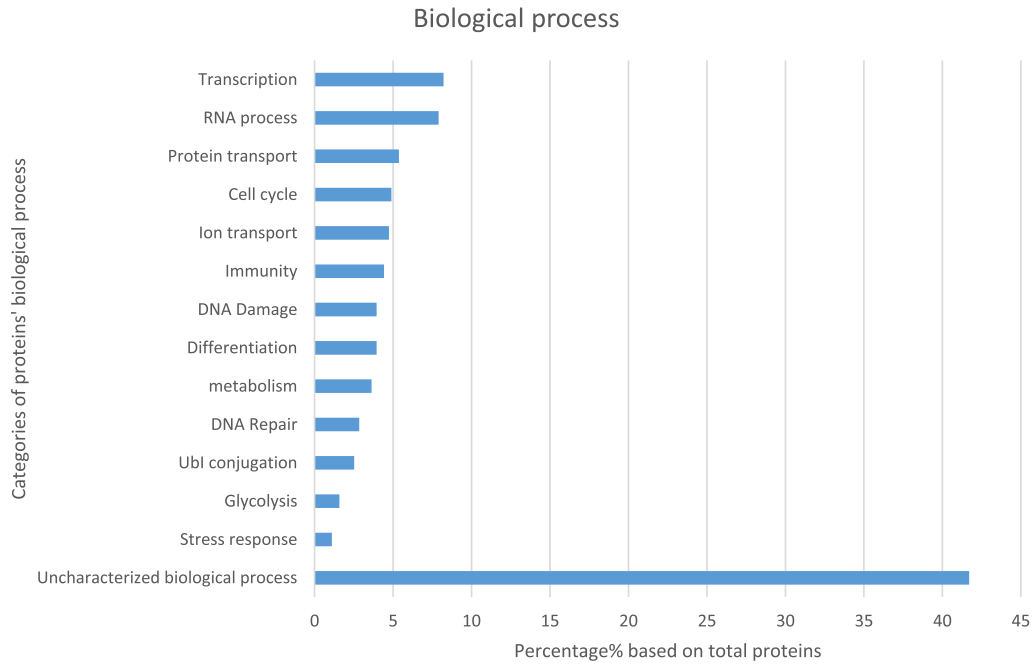
Fig. 2. Classification of donkey milk protein groups according to their biological process. The biological process of many proteins has not been characterized yet.
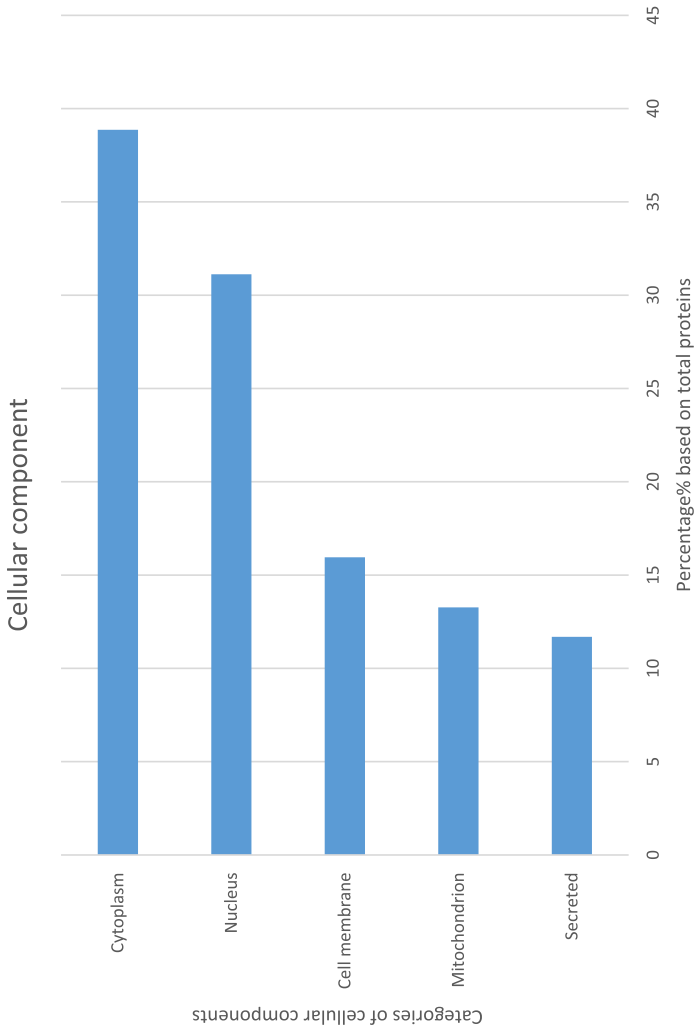
Fig. 3. Classification of donkey milk protein groups according to their cellular component.

2.2. Sample preparation

The samples' working aliquots were thawed on ice and then regenerated at room temperature. Next they were centrifuged at 4000xg for 1 h at 4 °C, for final fragmentation in three layers (lipid-, whey- and casein- layer). The flowchart of the strategy followed including the end-process for protein identification approaches used is represented previously [3]. From each sample the whey fraction was extracted and the protein content was determined the Bradford assay.

2.3. Peptide generation

For the peptide generation, 200 μg proteins of whey fraction was treated with 100 μL sample buffer solution (7 M urea, 2 M Thiourea, 4% CHAPS, 1% DTE), gently sonicated in a water bath for 30 minutes in room temperature. The sample is placed in an Amicon Ultra 0.5 Centrifugal Filter Device (Merck, Darmstadt, Germany) and 50 μL of iodoacetamide solution (55 mM) is added to the filter [4]. The solution remains at room temperature for 20 min in the dark, followed by centrifugation for 10 min at 13000xg at 4 °C. In the filter, 50 μL of Urea solution (8 M Urea, 0,1 M Tris HCl, pH 8,5) was added and the sample was centrifuged for 15 min at 13000xg at 4 °C. Then, 50 μL of 50 mM ammonium bicarbonate solution was added to the filter and the sample is centrifuged at 13000xg at 4 °C. Subsequently, trypsin solution 10% v/v (Roche Diagnostics, Basel, Swiss) in final concentration of 500 ng/μl is added to the sample, incubated at room temperature overnight at a ratio 1 μg per 50 μg protein . After incubation the sample is centrifuged at 13000xg at 4 °C for 10 min and 40 μL of 50 mM of ammonium bicarbonate solution is added and the sample is recentrifuged for 15 min at 13000xg at 4 °C. Finally, the filter discard, the tubes are placed in the speedvac for 2 h and then the tubes with the pellets of the digested sample stored at 4 °C for LC-MS/MS analysis.

2.4. 1-D nanoLC-MS/MS analysis

A modified protocol was followed in order to enhance identification rate of extracted peptides. LC-MS/MS analysis was performed exactly as described previously [5,6]. The pellets were dissolved in 100 μL of 0.1% formic acid in double-distilled water (phase A) for the LC-MS/MS analysis. The analysis of digested samples was made by using LTQ Orbitrap Elite coupled to Dionex 3000 nanoHPLC System (Thermo Scientific, Rockford, IL, USA). 6 μL of sample (approximately 5 μg of peptides) were loaded on a C-18 precolumn (100 m inner diameter x 2 cm; 100 Å, 3-m-bead-packed, Acclaim PepMap 100, Thermo Scientific). The procedure of the injection is automated. The mobile phase of the precolumn (A) was 0.1% formic acid in water. After 6 min of desalting the pre-column was removed online and the analytical C-18 column (75 m × 50 cm; 100Å, 2-m-bead-packed Acclaim PepMap RSLC, Thermo Scientific) took its place. The mobile phase A was 0.1% formic acid in water and the mobile phase B was 99% acetonitrile in water. The gradient elution curve was described as: 2.0% B (98.0% A) 10 min, 2.0-35.0% B (98.0-65.0% A) 325 min, 80.0% B (20.0% A) 10 min, 2.0% B (98.0% A) 10 minutes [5]. Samples were run at a constant 0,3μl/min flow rate.

The data were collected in data-dependent tandem mass spectometry mode using the standard top-20 method. The Orbitrap Elite mass spectrometer is connected with a nano-spray source. The ionization voltage settled at 2.20 kV and the capillary temperature is set at 250 °C. The instrument was operated in a positive ion mode and was measured at a resolving power of 60,000 with a maximum integration of 120 ms. The scans were performed in a higher energy collision dissociation (HCD) mode with a normalized collision energy of 36% and the full-scan data were acquired on the 250–1,250 m/z range. For HCD of parental ions, activation time of 0.1 ms was used. Ions with m/z fraction over 2 were subjected to MS/MS analysis. Measurements were performed using m/z 445.120025 as lock mass. The resulting MS/MS spectra were

acquired with 15,000 resolving power and a maximum injection time of 120 ms. Dynamic exclusion settings were set to repeat count 1, repeat duration 30 s, exclusion duration 120 s, and exclusion mass width 0.6 m/z (low) and 1.6 m/z (high). Identification criteria were performed using a precursor mass tolerance of 10 ppm and a 0.05 fragment mass tolerance. Trypsin was selected as the cleavage enzyme with a maximum of 2 missed cleavage points. A false-discovery rate threshold of 1% validated the peptide spectral matches (PSM). The minimum length of acceptable identified peptides was set as 6 amino acids.

The *.raw data files were analyzed using the Proteome Discoverer software (Thermo Scientific), using the Sequest search engine applying the *Mammalia* and the *Equus asinus* *.fasta databases. Both of them include reviewed entries. The repeatability between the 3 repetitions of the same sample is more than 85%, between the 6 samples analyzed the repeatability is 72%.

Ethics Statement

All samples included are commercially available and no animal experiments were performed in that study.

CRediT Author Statement

Stavros C. Proikakis: Investigation, Writing – original draft preparation; **Efterpi V. Bouroutzika:** Investigation; **Athanasios K. Anagnostopoulos:** Investigation, Methodology, Formal analysis; **George Th. Tsangaris:** Conceptualization, Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships which have or could be perceived to have influenced the work reported in this article.

Acknowledgments

Supplementary Materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.dib.2021.107507.

References

[1] V. Spada, P. Ferranti, L. Chianese, E. Salimei, F. Addeo, G. Picariello, Antibacterial potential of donkey's milk disclosed by untargeted proteomics, J. Proteom. 231 (2021) 104007, doi:10.1016/j.jprot.2020.104007.

[2] G. Theodoridis, A. Pechlivanis, N.S. Thomaidis, A. Spyros, C.A. Georgiou, T. Albanis, I. Skoufos, S. Kalogiannis, G.T. Tsangaris, A.S Stasinakis, I. Konstantinou, A. Triantafyllidis, K. Gkagkavouzis, A.S Kritikou, M.E Dasenaki, H. Gika, C. Virgiliou, D. Kodra, N. Nenadis, I. Sampsonidis, G. Arsenos, M. Halabalaki, E. Mikros, The FoodOmicsGR Ri Consortium OBO, FoodOmicsGR_RI. A consortium for comprehensive molecular characterisation of food products, Metabolites 11 (2) (2021) 74 2021, doi:10.3390/metabo11020074.

[3] A.K. Anagnostopoulos, A.I. Katsafadou, V. Pierros, E. Kontopodis, G.C. Fthenakis, G. Arsenos, S.C. Karkabounas, A. Tzora, I. Skoufos, G.Th. Tsangaris, Milk of Greek sheep and goat breeds; characterization by means of proteomics, J. Proteom. 147 (2016) 76–84, doi:10.1016/j.jprot.2016.04.008.

[4] J.R. Wiśniewski, K. Zettl, Datasets: Sensitivity and protein digestion course of proteomic filter aided sample preparation, Data Brief 26 (2019) 104530, doi:10.1016/j.dib.2019.104530.

[5] A.K. Anagnostopoulos, A.I. Katsafadou, V. Pierros, E. Kontopodis, G.C. Fthenakis, G. Arsenos, S.Ch Karkabounas, A. Tzora, I. Skoufos, G.T. Tsangaris, Dataset of milk whey proteins of three indigenous Greek sheep breeds, Data Brief 8 (2016) 877–880, doi:10.1016/j.dib.2016.06.040.

[6] A.K. Anagnostopoulos, D.J. Stravopodis, G.T. Tsangaris, Yield of 6000 proteins by 1D nLC-MS/MS without pre-fractionation, J. Chromatogr. B Analyt Technol. Biomed. Life Sci. 1047 (2017) 92–96, doi:10.1016/j.jchromb.2016.08.031.