IRSTI 34.15.29, 34.33.27 https://doi.org/10.26577/eb.2021.v89.i4.01

S.Sh. Atavliyeva, P.V. Tarlykov*

M. Republican State Enterprise «National Center for Biotechnology» under the Science Committee of Ministry of Education and Science of the Republic of Kazakhstan, Kazakhstan, Nur-Sultan
*e-mail: tarlykov@biocenter.kz

PALEOPROTEOMICS STUDIES OF ANCIENT *CAPRINAE*: A REVIEW

In the Neolithic era, people began to graze sheep and goats primarily due to easier access to meat, milk, and fleeces. Thus, in ancient times, Caprinae were the key animals in the development of early domestication and agriculture. The analysis of ancient proteins of Caprinae from paleontological and archeological materials reveals new data on their migration, complements research on the diet of ancient people, their culture, and habits. Here, we discuss paleoproteomics methods, such as matrix-assisted laser desorption/ionization with time-of-flight mass analyzer (MALDI-TOF MS) and liquid chromatography with tandem mass spectrometry (LC-MS / MS). We will also consider the most important discoveries in the field of the study of ancient sheep and in which direction the paleoproteomics of Caprinae will develop in the near future. In addition, general recommendations for analyzing data from ancient proteins are considered, for example, programs and requirements for databases. We will consider the main search algorithms in proteomics, as well as identify effective ones for identifying peptides and proteins. It also describes the commonly used ancient protein targets, and the basic principles of working with ancient samples. In addition, this review describes the main research conducted on ancient Caprinae of ancient proteins such as collagen, keratin, and milk proteins.

Key words: ancient protein, paleoproteomics, Caprinae.

С.Ш. Атавлиева, П.В. Тарлыков*

Қазақстан Республикасы Білім және ғылым министрлігінің Ғылым комитеті «Ұлттық биотехнология орталығы» республикалық мемлекеттік кәсіпорны, Қазақстан, Нұр-Сұлтан қ.
*e-mail: tarlykov@biocenter.kz

Ежелгі *Caprinae* палеопротеомдық зерттеулері: шолу

Неолит дәуірінде адамдар қой мен ешкілерді алдымен ет, сүт және жүнге оңай қол жеткізуге болатындықтан баға бастады. Осылайша, ежелгі уақытта Caprinae ерте қоныстану мен егіншіліктің дамуындағы негізгі жануарлардың бірі болған. Палеонтологиялық және археологиялық материалдар бойынша Caprinae ежелгі ақуыздарын талдау олардың көші-қоны туралы жаңа мәліметтер ашады, ежелгі адамдардың тамақтану рационы, мәдениеті мен әдеттерін зерттеуді толықтырады. Бұл мақалада палеопротеомика әдістерін талқылаймыз, сондай-ақ, матрицалық белсендірілген лазерлік десорбция/уақыт аралығының масс-анализаторымен иондану (MALDI-TOF MS) және тандем масс-спектрометриясы бар сұйық хроматография (LC-MS/MS). Сондай-ақ, ежелгі қойларды зерттеу саласындағы маңызды жаңалықтарды және caprinae палеопротеомикасы жақын арада қандай бағытта дамитынын қарастырамыз. Сонымен қатар, ежелгі ақуыздардың деректерін талдаудағы жалпы ұсыныстар қарастырылады, мысалы, бағдарламалар, алгоритмдер және мәліметтер базасына қойылатын талаптар. Сонымен қатар, протеомикадағы негізгі іздеу алгоритмдерін сипаттап, пептидтер мен ақуыздарды талдау үшін олардың тиімдісін анықтаймыз. Бұл мақалада біз палеопротеомиканың әдістерін, ежелгі таралған ақуыз субстраттарын және ежелгі үлгілермен жұмыс істеудің негізгі принциптерін талқылаймыз. Сонымен қатар, бұл шолу коллаген, кератин және сүт ақуыздары сияқты ежелгі Caprinae ақуыздарының негізгі зерттеулерін сипаттайды.

Түйін сөздер: ежелгі ақуыздар, палеопротеомика, Caprinae.

С.Ш. Атавлиева, П.В. Тарлыков*
Республиканское государственное предприятие «Национальный центр биотехнологии»
Комитета науки Министерства образования и науки Республики Казахстан, Казахстан, г. Нур-Султан
*e-mail: tarlykov@biocenter.kz

Палеопротеомные исследования древних *Caprinae*: обзор

В эпоху неолита люди начали пасти овец и коз, в первую очередь, из-за более легкого доступа к мясу, молоку и шерсти. Таким образом, в древние времена Caprinae были ключевыми животными в развитии раннего одомашнивания и земледелия. Анализ древних белков Caprinae по палеонтологическим и археологическим материалам открывает новые данные об их миграции, дополняет исследования рациона питания древних людей, их культуры и привычек. Здесь мы обсуждаем методы палеопротеомики, такие как матрично-активированная лазерная десорбция/ионизация с времяпролетным масс-анализатором (MALDI-TOF MS) и жидкостная хроматография с тандемной масс-спектрометрией (LC-MS/MS). Также рассмотрим важнейшие открытия в области изучения древних овец и в каком направлении палеопротеомика Caprinae будет развиваться в ближайшее время. Дополнительно рассмотрены общие рекомендации при анализе данных древних белков, например программы и требования к базам данных. Кроме того, рассмотрим основные в протеомике поисковые алгоритмы, а также выявим эффективные из них для идентификации пептидов и белков. Также описаны древние субстраты для выделения белка и основные принципы работы с древними образцами. В этом обзоре описываются основные исследования белков древних Caprinae, таких как коллаген, кератин и молочные белки.

Ключевые слова: древние белки, палеопротеомика, Caprinae

Introduction

Research into ancient biomolecules, especially DNA and proteins, has changed our understanding of evolutionary history, animal domestication, and phylogeny. Previously, discoveries were carried out based on archaeological excavations with the analysis of living organisms and the observation of phenotypic features in the fossils. Studies of ancient biomolecules supplement and open up new knowledge, providing information on phylogeny, ancient migration, and the evolution of species.

Ancient DNA, as an object of research, has made a significant contribution to the development of archeology, complementing phenotypic research. Despite ongoing analyzes of evolutionary processes with a high level of confidence, nucleic acids are fragmented over time into increasingly shorter sequences. In this case, proteins and lipids are prioritized in studies with older samples in geographic areas that are less favorable for DNA preservation. It is worth noting that ancient proteins also undergo fragmentation over time, but compared to nucleic acids, they do this more slowly [1, 2].

Over the past two decades, with the advent of highly sensitive mass spectrometry, paleoproteomics has become more and more in demand in the fields of archeology and evolutionary biology. Thus, researchers are focusing not only on ancient DNA but also on ancient proteins. The coverage of tissues and substrates for protein extraction is quite wide and includes bones, dentin, enamel, tartar, horns, eggshells, skin and soft tissues, various food residues, and ceramics [2]. This difference in objects in ancient samples is interesting for the analysis of complex mixtures of proteins produced by an individual organism (proteome) or a group of organisms (metaproteome) [3, 1].

In this review, we describe the main research methods used to study ancient proteins, substrates for protein isolation, characterization of ancient proteins, and summarize data processing and data interpretation. Then we will consider the most important discoveries in the field of the study of ancient sheep, and in which direction the paleoproteomics of *Caprinae* will develop soon.

Methods in paleoproteomics

The study of ancient proteins using mass spectrometry dates back to 2000 when osteocalcin from ancient bone was discovered [4]. Until this time, attempts to sequence proteins have been unsuccessful. The Edman sequencing method available at that time turned out to be unsuitable for ancient proteins since this method required samples with a high concentration, unmodified and purified – such conditions are incompatible with ancient biomolecules proteins [5].

The application of mass spectrometry in proteomics is quite wide, the principle is that molecules are ionized and identified by their mass-to-charge

ratio (m/z). As a result, mass spectra are obtained in the form of graphs of the relative content of ions in the sample to their m/z values [6]. Currently, paleoproteomic studies are mainly carried out using two mass spectrometry methods: matrix-based laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) and liquid chromatography with tandem mass spectrometry (LC-MS/MS). These methods are based on the presence and detection of single amino acid polymorphisms (SAP) between homologous protein sequences of different genera, species, or populations. Variations in the protein sequence occur at the genome level from single nucleotide polymorphisms (SNPs) in the gene encoding the protein. This SAP becomes the key for phylogenetic analysis of ancient proteins. Thus, the relationship between the proteome and the genome is traced in evolutionary research in the taxonomic analysis [2].

Figure 1 shows the main differences between the MALDI-TOF MS and LC-MS/MS methods. Thus, using MALDI-TOF MS, one can only get an idea of the total mass of individual peptides in the analyte, while LC-MS/MS determines the exact amino acid sequence of the peptides. Both methods have their limitations and application features, which will be discussed below [2].

MALDI-TOF MS

The MALDI-TOF MS method is widely used for peptide mass fingerprinting (PMF), based on species identification by comparing the peptide profile of an unknown sample with the masses of known peptides. The peptide profile results from enzymatic cleavage by proteases such as trypsin [7].

The initial application of this approach to ancient fauna fossils was phylogenetic identification based on collagen peptide mass fingerprinting (CPMF) [8]. The method, also called ZooMS (Zooarcheology by Mass Spectrometry), aims to identify the trypsin-digested collagen protein or other proteins, using MALDI-TOF MS to measure the mass-to-charge ratio (m/z). This method is generally similar to DNA fingerprinting, and trypsin, like a restriction enzyme, cuts molecules into fragments. Therefore, CPMF is not suitable for examining tissues that do not contain COL1 or are present in small amounts. In skin, bones, dentin, and horns, COL1 is a key protein and can be preserved in fossils about 600,000 years ago. The protein consists of two alpha-1 chains of type I collagen (COL1a1) and one alpha-2 chain of collagen type I collagen (COL1a2), which twist on top of each other to form a triple helix [2, 9]. By measuring the m/z of individual COL1 peptides, certain patterns of fragments can be obtained, which are used as a comparison, as peptide markers. The masses of homologous peptide sequences in species and genera may differ if they contain one, two, or several SAPs [2, 9].

ZooMS is presented as a simple method for identifying taxa that are difficult to identify by morphological characteristics. So, using this method, identification was carried out between goats (*Capra sp.*) and sheep (*Ovis sp.*), which is difficult to accurately determine during archaeological excavations [10]. It is worth noting that other proteins (keratins, caseins) have also been studied, which makes it possible to apply PMF to other ancient tissues that do not contain collagen. New peptides detected by MALDI-TOF MS can be confirmed by LC-MS/MS.

LC-MS/MS

Tandem mass spectrometry allows the sequencing of whole and mixed metaproteomes. Consequently, this approach to the study of paleoproteomics is more universal and is used in a wide range of different tissues of ancient samples. For this method, one may not know in advance about the protein sequence, thus it is possible to detect new amino acid substitutions. However, it should be kept in mind that ancient proteomes are not numerous and when loading a sample into the device, only a part is identified, therefore the device settings should be adapted more towards the sensitivity of the device [1, 2].

Shotgun proteomics was originally based on the study of the COL1 protein, but over time the spectrum expanded to include whole proteomes. This method made a significant contribution to the study of extinct organisms, amino acid sequences were obtained based on which a phylogenetic reconstruction of evolutionary relationships with other extinct and currently existing organisms was carried out [11, 12].

Thus, LC-MS/MS also opens up the possibility of phylogenetic analysis of species whose DNA has not been preserved, and ancient proteins are available for study. Nowadays, towards the analysis of the amino acid sequence of ancient proteomes, shotgun proteomics expands the possibilities in the study of various modifications of proteins *in vivo* [2].

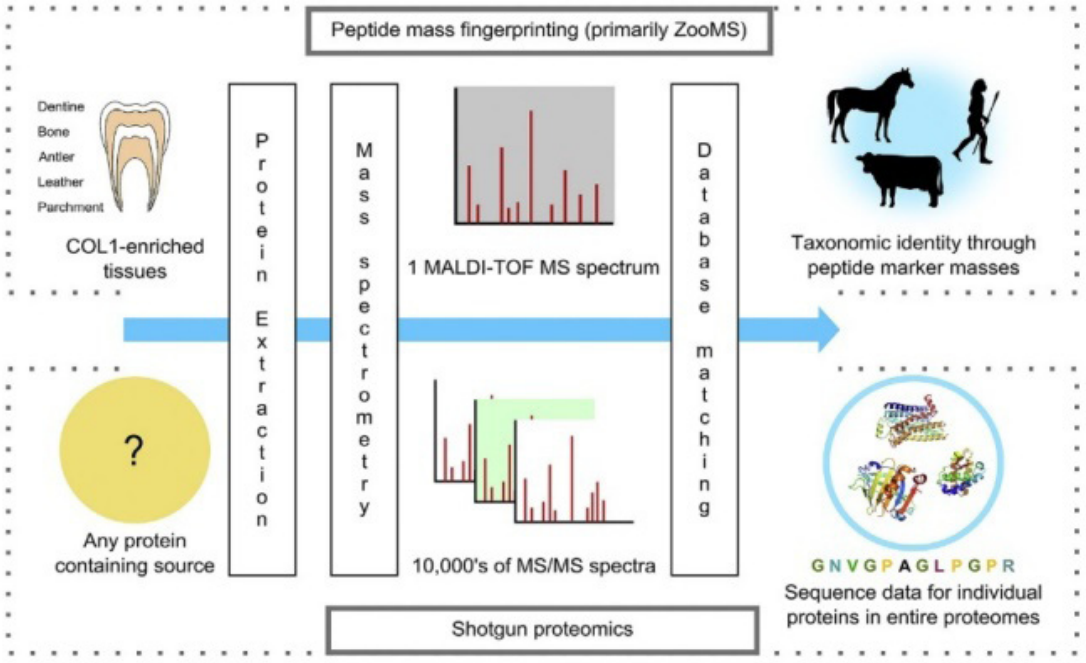

Figure 1 – Differences between MALDI-TOF MS (PMF) and LC-MS/MS (shotgun proteomics) workflows applied to study ancient proteins [2]

Types of biomaterial

Similar to the study of ancient DNA, the main protein materials in ancient samples are bones and teeth. Bone proteome studies have become common practice in the detection of collagen type 1 (COL1) in the case of fingerprinting of ZooMS collagen peptides [1, 2]. In terms of the qualitative composition of proteins, dentin and bone proteomes are similar, since they have a common origin from ectomesenchyme [13]. Thus, ancient dentin and bone samples contain the dominant amount of the COL1 protein [14]. A small fraction of their proteome is represented by non-collagen proteins (NCPs), such as albumin and biglycan, difficult to detect in mass spectrometry experiments. However, they contain more variations in the SAP sequence compared to collagens, thus being of phylogenetic interest [2].

To date, few studies of the enamel proteome have been carried out; it is worth noting that the studied archaeological samples are similar in composition to the enamel of modern samples, thus, it should be assumed that the enamel is well preserved over time and possibly protected from diagenetic effects [15, 16]. The enamel proteome differs significantly from bone and dentin, which contains specific proteins: amelogenins (AMELX and AMELY), ameloblastin (AMBN), amelotin (AMTN), emelin (ENAM) and odontogenic ameloblast-associated protein (ODAM), and matrix proteases (MMP20 and KLK4). Of particular interest is the protein amelogenin, whose gene is located on the X- (AMELX) and Y-chromosome (AMELY) [17]. Thus, identification of AMELY peptides using proteomic methods allows the determination of male sex (XY), and the absence of an amino acid sequence may indicate that the sample belongs to a female body (XX). Such gender identification can be applied not only to archaeological people but also to fauna samples, which is an alternative to DNA-based sex determination [16].

Recently, paleontological research has relied on alternative sources of protein, such as cultural heritage materials or ancient tooth tartar [6, 19]. A large amount of protein is present in the mummified remains of humans and animals, as well as garments made from animal skin [20].

Limitations and features of work with ancient proteins

Due to its high sensitivity, mass spectrometry has become a key method for the analysis of ancient proteins, despite this, standard methods of protein extraction must be modified and several recommen-

dations must be followed to effectively preserve the ancient protein at every stage of preparation [2].

When forming a collection of samples, it should be borne in mind that some types of biomaterial better preserve endogenous proteins; for example, the mineralized specimens mentioned above (bone, dental plaque, and eggshell). After choosing a sample for research, it is imperative to conduct trial experiments with a modern representative of an ancient organism and, if possible, with artificial diagenesis of the sample or less valuable ancient ones [2]. Additionally, MALDI-TOF MS can screen peptides for the degree of preservation of an ancient protein, especially before expensive LC-MS/MS analysis. Since this method is reliable for predicting the survival of proteins in fossils [21].

Contamination of ancient proteins can occur at any stage of research from excavation to protein analysis, it is worth taking measures to reduce the risks of contamination with modern samples and cross-contamination between the ancients by analogy with ancient DNA. General guidelines should be followed, such as isolating work areas and separate protective clothing in each area, including negative controls, clean surfaces and equipment, and avoid reusing consumables. Additionally, when analyzing LC-MS/MS, it is necessary to rinse the LC column between each sample, i.e. skip blanks as remaining peptides on the column may contaminate subsequent samples. For the same reason, it is recommended that valuable and older samples should be injected into the column first, to avoid false-positive results due to subsequent samples with a higher concentration [2].

Ancient proteins are usually fragmented, most often due to accidental non-enzymatic cleavage of the peptide backbone at the carboxyl side of asparagine (Asn) and glutamine (Gln). Also, in studies of ancient proteins, deamidation of asparagine (Asn) and glutamine (Gln) has been observed to form aspartic acid (Asp) and glutamic acid (Glu), respectively [1, 22]. Glutamine and asparagine are abundant in most proteins, and this wide availability of both amino acids is key when making comparisons of damage between proteins. The study of deamidation in ancient proteins is available as it can be quantified using both MALDI-TOF MS and LC-MS/MS. Thus, Welker et al. in the study of samples of the late and middle Pleistocene noticed that the deamidation of glutamine in NCP was significantly higher, almost 100% than in endogenous collagens from the same sample [23, 24]. This distinction has been proposed to be used as a marker between endogenous and contaminating NCPs [2]. This degradation at the amino acid level is a natural biomolecular marker of sample aging and leads to a mass shift of +0.984 Da [25].

The analysis of the spectra should take into account that the modification of amino acids leads to a change in the total mass, which may be equal to the mass of another amino acid. Such a problem arose in the study of the beta-lactoglobulin protein, it is known that the difference between *Bovinae* (cattle, yak, and buffalo) and *Caprinae* (sheep and goats) lies in the amino acid residue, in bulls – aspartic acid, in sheep – asparagine and in goats – lysine [26]. However, deamidation of asparagine results in its conversion to aspartic acid, so the unmodified *Bovinae* residue (D) and the deamidated sheep residue (de. N) in this position will show the same mass. In this case, the difference in species will be possible only in the absence of modifications [23, 26].

In addition to deamidation, another diagenetic damage to ancient proteins has been seen. For example, oxidative damage to tryptophan and lysine has been observed in mammoth bone with the formation of kynurenine and aminoadipic acid, respectively. In turn, carbonylation of arginine and lysine prevents trypsinolysis of the ancient protein into peptides, which reduces the quality of sample identification [27]. Interestingly, diagenesis can also have a stabilizing effect on proteins, as ancient mineral-coated specimens have been shown to retain proteins better [28].

Thus, knowledge of the factors and conditions of biomaterial storage helps in the study of ancient samples and diagenetic modifications of proteins require more detailed study.

Data analysis

The identification of peptides and proteins with LC-MS/MS analysis data is carried out using ready-made search algorithms Mascot or Sequest, which are the main ones in proteomics. However, the Byonic and Andromeda, PEAKS programs are more suitable for paleoproteomics, since their algorithms allow identifying unknown modifications and highly degraded protein samples [1, 29]. Sequencing of ancient proteins based on mass spectra should be based on at least two razor and unique peptides covering different regions of the sequence. Razor or shared peptides are peptides that belong to the group of proteins with the largest number of identified, shared peptides. Unique or proteotypic peptide sequences belong to only one protein [30].

Identification algorithms match the masses of the product ions against spectra from the database.

When identifying peptides, the choice of databases should not be limited only to the species of interest, it is also necessary to include databases of microbes and possible contaminants. Since, the composition of the proteome of ancient samples changes over time, becoming contaminated with a mixture of various bacteria, fungi and during excavation with human proteins. Consequently, the protein sequences of type I collagen and other human proteins, animal skin proteins are considered contaminants and should be included in the research database. It should be noted that the small volume of databases on the number of peptides can lead to incorrect identification of proteins and taxonomic definition of the type of organism. For example, the Swiss-Prot database includes almost complete proteomes of model organisms such as mice (*Mus musculus*) and human (*Homo sapiens*), but only partially proteomes of other taxa, such as sheep (*Ovis aries*), goat (*Capra hircus*), cow (*Bos taurus*) and pig (*Sus scrofa*). Difficulties can arise when analyzing metaproteomes isolated from a heterogeneous material, for example, the remains of ceramics or dental calculus. Thus, there is a database error when identifying several species from the same sample, and extensive databases should be consulted to avoid incorrect taxa [3].

After data generation, several additional analyzes are performed to further validate and validate the results. For example, public code is used to evaluate deamidation, and contaminating proteins must be filtered out [31].

Exploring ancient *Caprinae* through paleoproteomics

Studies of ancient biomolecules of *Caprinae* have revealed the history of their domestication, migration routes, and the way of life of ancient people. To date, in the paleoproteomics of *Caprinae*, studies have been carried out on the peptide difference between sheep and goats, ancient textiles, the diet of ancient people, and objects of art for the content of animal raw materials have been studied. Here we review some of the basic research.

The domestication of animals such as cattle, sheep, goats, and horses was an important step that changed the strategy of the ancient people and the transition from hunting to agriculture. It is known that the domestication of *Caprinae* dates back to 11,000-9,000 years ago was carried out in the territory from eastern Anatolia to the Zagros mountains in Iran and Iraq [32]. The spread of sheep and goats from the center of domestication was part of the Neolithic culture, so livestock was dispersed across Eurasia and Africa [33]. Initially, livestock was used for the production of meat, skins, and bones, and then for wider use, for riding, transporting, or cultivating the soil, they were also a source of milk and wool [32, 34].

Methods of paleoproteomics expand the range of studies of ancient samples, so it can be used to determine which species the bones belong to, especially when the remains are highly fragmented or insufficient for analysis of aDNA. For identification, ancient proteins are separated into peptides, followed by analysis on a mass spectrometer. The treated peptides have m/z values that form amino acid sequence fragmentation patterns specific to each species [35].

Collagen

During archaeological excavations, sheep and goats are identified as a single whole, however, they, as well as cattle, did not always migrate together. There are differences in food and water needs, grazing patterns, travel speeds, all of which point to the need for an individual story [36]. Thus, to distinguish between the remains of a domestic sheep (*Ovis aries*) and a domestic goat (*Capra hircus*), one of the most common methods used in paleoproteomics was applied, namely, ZooMS. Anatomically, it is difficult for archaeologists to distinguish small artiodactyls from bone remains, but ZooMS has shown its effectiveness in various parts of the world [10, 36, 37]. Research has been done using collagen protein as one of the strongest and most affordable ancient proteins.

Originally Buckley et al. described the differences between sheep and goats for one collagen type 1 protein peptide (COL1). The *de novo* sequenced peptide consists of 33 amino acids and differs between species at two positions. Differences in the amino acid sequence led to unequal m/z of the peptides, for example, in sheep, m/z of the peptide was 3033.3±0.2, and in goats, m/z was 3093.3±0.1 (Table 1). The proposed marker was tested on archaeological samples of *Caprinae* from the Neolithic settlement Domuztepe in Turkey [10].

Interestingly, in archaeological samples, the mass of the marker peptide was increased by 1 Da, which is associated with the PTM deamidation of Glu (Q). Thus, the m/z for the ancient samples of the sheep became equal to – 3034 Da, and in the goat, respectively – 3094 Da (Figure 2). Also, in the differentiated peptide was found another PTM known as hydroxyproline. In total five hydroxylated Pro (O) residues were observed in both modern and ancient peptides [10]. Later, the same research group

expanded the collagen mass fingerprinting method by additionally adding modern species of other animals (deer, gazelle) to domestic animals (sheep, goats, pigs, and cattle), as well as humans [37]. Another group of researchers found another additional peptide (Table 1) differences between the *Caprinae* protein COL1A1, located at positions 921-936. Interestingly, this peptide is common and does not differ in mass for *Capra hircus, Bos Taurus,* and *Homo sapiens* [38].

Table 1 – Amino acid sequences of marker peptides of type I collagen between Caprinae species. The underlined amino acid residues differ from honey by species [10, 38]

Species	Peptide amino acid sequence	m/z
Ovis aries	GPSGEOGTAGPOGTOGPQG**L**LG**A**OGFLGLOGSR	3033±0, 2
Capra hircus	GPSGEOGTAGPOGTOGPQG**F**LG**P**OGFLGLOGSR	3093,3±0,1
Ovis aries	**A**GEVGPPGPPGPAGEK	709,3
Capra hircus	**P**GEVGPPGPPGPAGEK	722,3

Species confusion when identifying sheep can arise not only with goats but also with wild species, such as medium-sized cattle (impala, antelope). In this case, the use of ZooMS is of great importance, since the data on the presence of domesticated animals or wild animals at ancient sites change the history of the distribution of the Neolithic around the world with its later appearance in some parts of the world. In a recent study of the Leopard Cave in South Africa, previously anatomically defined samples of the domestication of sheep and goats turned out to be wild antelope species. The peptide sequences were compared to 20th-century museum collectibles of antelope and impala. A total of three new taxonomic markers in the alpha 2 chain of type I collagen have been identified to distinguish between wild antelope and impala. This study showed that *Caprinae* appeared in the Leopard Cave 1,500 years later than previously thought [29].

A similar study, but with other species of wild animals (*Sylvicapra grimmia* – gray duiker, *Pelea capreolus* – gray rhebok, *Antidorcas marsupialis* – springbok), anatomically the same size as sheep, was carried out by the research group Coutu et al. Using paleoproteomics methods, a marker peptide was determined that distinguishes wild bovids from sheep; this peptide m/z 1532 is absent in the collagen sequence COL1A2 position No. 889–906 *Ovis aries*. It has also been confirmed that a specimen dating from about 2000 BC belongs to a species of domestic sheep, which is the earliest evidence of domesticated animals in southern Africa [31].

Mass spectrometry techniques were applied to art from the 14th century, as *Ovis aries* collagen peptides were found on a mural by Ambrogio Lorenzetti. In addition, egg whites from chicken, duck, and cow glue were also identified, all of which indicate the use of animal proteins as a conservation product. Thus, paleoproteomics applies to cultural heritage sites without destructive action, for a better understanding of history and the use of materials in antiquity [22].

Keratin

Paleoproteomics also made an important contribution to the study of cultural heritage, which made it possible to determine the biological origin of many keratin-containing materials using the example of ancient textiles. The main proteins that make up ancient clothing are keratins. They constitute an extensive family of fibrillar proteins, which, due to a large number of disulfide bonds, have high strength. Keratins are divided into two classes: alpha-keratins are helical proteins found in mammals, birds, and reptiles, and lamellar beta-keratins, which are found only in birds and reptiles [7, 39]. According to the mammalian nomenclature, keratins are divided into 2 types: type I from K33 to K40 and type II from K81 to K87 [6, 35].

Ancient people used sheep and goat wool to create fabrics and fur clothing [6]. As is customary, the morphological study of ancient tissues is carried out using microscopic methods, so the families of animals are determined with high accuracy. However, these methods are difficult to apply to severely damaged tissue samples or tissue samples made from closely related animal species [35]. Keratins are highly conserved proteins and identification by mass spectrometry between closely related species is difficult. However, a keratin marker peptide to distinguish between sheep and goats has been proposed K33, which differs by only one amino acid in the sequence. Keratin K33 usually exists in two isoforms named K33a and K33b [25, 40].

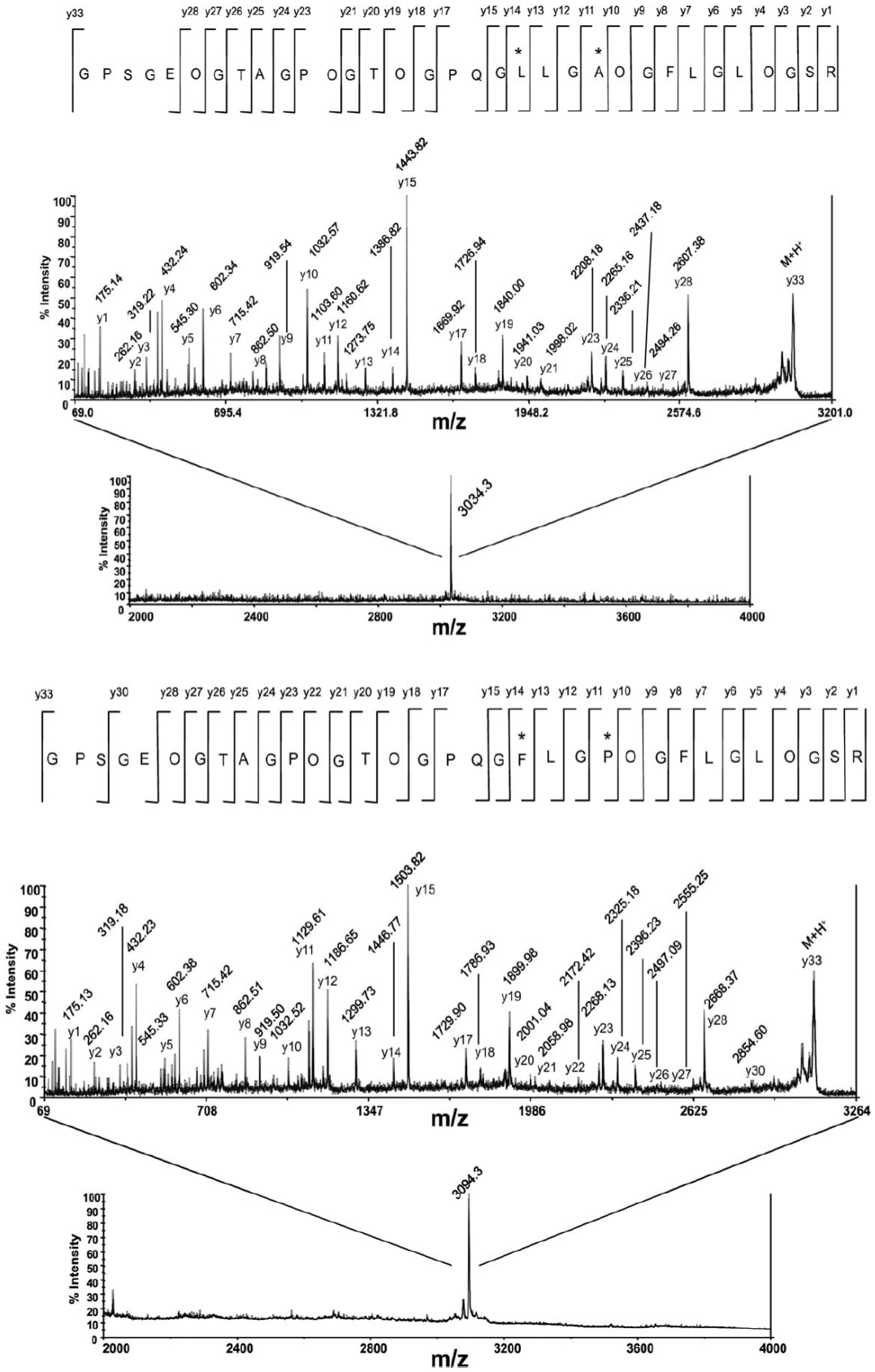

Figure 2 – The results of de novo sequencing of a collagen-peptide ion spectrum showing the marker m/z values and distinction between the archaeological sheep (top) and goat (bottom) [10]

Thus, keratin-containing archaeological remains were studied using PMF and MS/MS. One of the earliest was the study of the ancient clothing of Ötzi, it was found that the outer clothing was made of sheep, and the shoes were made of cattle, other parts of the clothing consisted of deer and goat skins [41].

The study of ancient fabrics is mainly focused on the monuments of the Bronze Age, the Iron Age, or places of trade routes. Since cattle breeding, the production of wool and fabric flourished at that time, as a result, cultural exchange through the Great Silk Road [25]. One of these ancient sites in China in the Keriya Valley showed that ancient people preferred to use more sheep (57.8%) in a herd than goats (16.5%), while woven products were made mainly from sheep's wool, and goats – skin [35]. However, unlike other materials, tissues are sensitive and poorly preserved, degrade faster, and are prone to deamidation. It is known that the deamidation of glutamine (Q) and asparagine (N) is a marker of the degradation and aging of ancient proteins. Using the example of wool sheep wool, 8 alpha-keratin peptides were found with deamidation indicating degradation of the samples, of which five are more stable (m/z 1487.74, 1504.77, 1625.84, 1834.97, and 2665.30), and three are possibly poorly identified due to strong degradation. Thus, the study of textiles using paleoproteomics can provide reliable and in-depth information about resources, production technologies, trade, and the culture of antiquity over time [25].

The ancient proteins of keratins are also examined in other organic materials such as horns, hooves, and hair [7]. Interestingly, they, like wool, consist of the same type of keratin and have the same amino acid sequences, but differ in PMF. This may be due to differences in protein expression. For example, a peptide with m/z 2519.29 (YSSQLAQMQGLIGNVEAQLAEIR) found in the horns of yak, cattle, and sheep is absent in fibers [40]. Despite this, scientists were able to find marker peptides for alpha-keratin to determine the species. Thus, the peptide m/z 2665 is diagnostic for sheep, a similar peptide for goats with m/z 2692, for a horse – m/z 2563, for a cow – m/z 2577 [7]. This study shows that PMF can also be used for keratinized tissues, especially for remains with severe degradation, when the structure of the horn is not preserved.

Milk proteins

The study of dairy products of animal husbandry is significant in the history of the development of the diet of ancient people. In particular, the identification of the animal species used for milk production allows the tracking of shifts in the exploitation of cattle, sheep, and goats. The remains of food in the tomb M27 at the Subaeisi site in China 500-300 years BC were studied for the presence of dairy products. Were identified four sequences of peptides αs1-casein (FVVAPFPEVFR, YLGYLEQLLR, FFVAPFPEVFGK) and β-casein (DMPIQAFLLYQEPVLGPVR) belonging to sheep, goats, cattle, and buffalo [34]. Proteomic approaches based on mass spectrometry, in particular, have been used to clearly distinguish between animal species for milk proteins. In another study, the diet of ancient people of the Late Bronze Age was studied, the remains of food were found in the form of tartar. The mass spectrometric analysis identified the peptide sequences of the αs1-casein protein (FVVAPFPEVFR, FFVAPFPEVFGK) and the β-lactoglobulin protein (TPEVD(D/N/K) EALEKFDK), which contains the genus-specific amino acid: D – *Bos*; N – *Ovis*, K – *Capra* [19].

In a study of the tombs of ancient Egypt, a dairy product similar to cheese was discovered, this sample became the oldest hard cheese (3200 years ago) found during excavations. The MS/MS method confirmed the presence of milk proteins in the ancient sample, a total of nine *Bovidae* peptides (cow, sheep, goat, buffalo) were found. Most, that is, six peptides belonged to casein (αs1-, β- and κ-), and the remaining three peptides to lysozyme and serum albumin [20].

In addition to casein, β-lactoglobulin (BLG) has been found in ancient dental calculus samples from British Neolithic sites. All peptides found were attributed to *Bovidae*, as the authors suggest, the Neolithic settlement used several species of animals (cows, goats, and sheep) for milk production [42].

Palaeoproteomic studies of milk proteins have contributed to the nutritional characteristics of ancient humans with the identification of the origin of raw materials, which helps to better understand the exploitation of domestic animals in the past.

Conclusion

To date, paleoproteomics in combination with other methods has made it possible to obtain reliable information about antiquity, since research covers continents, different eras, and the main groups of organisms. This was facilitated by the widespread use of mass spectrometry methods, which are distinguished by their high sensitivity, as well as the development of data processing algorithms. The main types of ancient protein substrates, the features of their isolation and preservation have been described and studied. Studies of ancient *Caprinae* proteins

have added knowledge about the distribution of the Neolithic around the world, so with the help of peptide mass fingerprinting, sheep, goats, and wild representatives of the fauna that are anatomically similar to them were differentiated. Also studied are the ancient proteins of keratin contained in clothing, fabrics at ancient sites. The transition of ancient people from primary livestock products to secondary ones can be traced to the use of dairy products. The protein composition of food remains was studied and the domestic animals used by ancient people for the production of dairy products were identified. However, today there are limitations for paleoproteomic methods such as a requirement for complete databases of amino acid sequences. Consequently, the main task in the study of ancient proteins is the replenishment of the database, especially with unique ancient proteins.

Financing

This research was funded by the Science Committee of the Ministry of Education and Science of the Republic of Kazakhstan (Grant No. AP08955854).

References

1 Cappellini E., Prohaska A., Racimo F., et al. Ancient Biomolecules and Evolutionary Inference // Annual Review Biochem. – 2018. – Vol. 87. – P. 1029-1060.

2 Welker F. Palaeoproteomics for human evolution studies // Quat. Science Rev. – 2018. – Vol. 90. – P. 137-147.

3 Hendy J., Welker F., Demarchi B., Speller C., Warinner C., Collins M.J. A guide to ancient protein studies // Nat. Ecol. Evol. – 2018. – Vol. 2, No 5. – P.791-799.

4 Huq N.L., Tseng A., Chapman G.E. Partial amino acid sequence of osteocalcin from an extinct species of ratite bird // Biochem. Int. – 1990. – Vol.21, No 3. – P. 491–496.

5 Robbins L.L., Muyzer G., Brew K., Macromolecules from living and fossil biominerals. In: Engel, M.H., Macko, S.A. (Eds.), Organic Geochemistry: Principles and Applications. – 1993. – Vol. 11. – P. 799-816.

6 Solazzo C. Characterizing historical textiles and clothing with proteomics // Conserv. Patrim. – 2019. – Vol. 31. – P. 97-114.

7 O'Connor S., Solazzo C., Collins M. Advances in identifying archaeological traces of horn and other keratinous hard tissues // Stud. Conserv. – 2015. – Vol. 60. – P. 393-417.

8 Buckley M., Matthew Collins M. et al. Species identification by analysis of bone collagen using matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry Michael // Rapid. Commun. Mass Spectrom. – 2009. – Vol. 23. – P. 3843–3854.

9 van Doorn N.L., Hollund H., Collins M.J. A novel and non-destructive approach for ZooMS analysis: Ammonium bicarbonate buffer extraction // Archaeol. Anthropol. Sci. – 2011. – Vol. 3. – P. 281-289.

10 Buckley M., Whitcher Kansa S., Howard S., Campbell S., Thomas-Oates J., Collins M. Distinguishing between archaeological sheep and goat bones using a single collagen peptide // J. Archaeol. Sci. – 2010. – Vol. 37. – P. 13-20.

11 Buckley M., Fariña R.A., Lawless C., Tambusso P.S., Varela L. et al. Collagen Sequence Analysis of the Extinct Giant Ground Sloths Lestodon and Megatherium // PLOS ONE. – 2015. – Vol. 10. e0144793.

12 Buckley M. Ancient collagen reveals evolutionary history of the endemic South American "ungulates." // Proc. R. Soc. – 2015. – Vol. 282. – P. 1-9.

13 Salmon C.R., Giorgett A.P.O., Paes Leme A.F., Domingues R.R., Kolli T.N., Foster B.L., Nociti Jr., F.H., Microproteome of dentoalveolar tissues // Bone. – 2017. – Vol. 101. – P. 219-229.

14 Salmon C.R., Tomazela D.M., Ruiz K.G.S., Foster B.L., Paes Leme A.F. et al. Proteomic analysis of human dental cementum and alveolar bone // J. Proteom. – 2013. – Vol. 91. –P. 544–555.

15 Porto I. M., Laure H.J., Tykot R.H., de Sousa F.B., Rosa J.C., Gerlach R.F. Recovery and identification of mature enamel proteins in ancient teeth // Eur. J. Oral Sci. – 2011. – Vol. 119. – P. 83-87.

16 Castiblanco G.A., Rutishauser D., Ilag, L.L., Martignon S., Castellanos J.E., Mejía W. Identification of proteins from human permanent erupted enamel // Eur. J. Oral Sci. – 2015. – Vol. 123. – P. 390-395.

17 Stewart N. A., Gerlach R.F., Gowland R.L., Gron K.J., Montgomery J. Sex determination of human remains from peptides in tooth enamel // PNAS. – 2017. – Vol. 114. – P. 13649–13654.

18 Stewart N.A., Molina G.F., Issa J.P.M., Yates N.A., Sosovicka M., Vieira A.R., Line S.R.P., Montgomery J., Gerlach R.F. The identification of peptides by nanoLC-MS/MS from human surface tooth enamel following a simple acid etch extraction // RSC Adv. – 2016. – Vol. 6. –P. 61673-61679.

19 Jeong C., Wilkin S., Amgalantugs T. et al. Bronze Age population dynamics and the rise of dairy pastoralism on the eastern Eurasian steppe // Proc. Natl. Acad. Sci. USA. – 2018. – Vol. 115. – P. 11248-11255.

20 Greco E., El-Aguizy O., Ali M.F. et al. Proteomic Analyses on an Ancient Egyptian Cheese and Biomolecular Evidence of Brucellosis // Anal Chem. – 2018. – Vol. 90. – P. 9673-9676.

21 Presslee S., Penkman K., Fischer R. et al. Assessment of different screening methods for selecting palaeontological bone samples for peptide sequencing // Journal Proteomics. – 2021. – Vol. 230.- P. 103986.

22 Mackie M., Rüther P., Samodova D., et al. Palaeoproteomic Profiling of Conservation Layers on a 14th Century Italian Wall Painting. // Angew. Chemie. – Int. Ed. – 2018. – Vol. 57. – P. 7369-7374.

23 Welker F., Hajdinjak M., Talamo S., Jaouen K., Dannemann M., David F., Julien M., Meyer M., Kelso J., Barnes I., Brace S., Kamminga P., Fischer R., Kessler B.M., Stewart J.R., Paabo S., Collins M.J., Hublin J.-J., Palaeoproteomic evidence identifies archaic hominins associated with the Chatelperronian at the Grotte du Renne // Proc. Natl. Acad. Sci. U. S. A. – 2016. – Vol. 113. – P. 11162-11167.

24 Welker F., Soressi M.A., Roussel M., van Riemsdijk I., Hublin J.-J., Collins, M.J.. Variations in glutamine deamidation for a Chatelperronian bone assemblage as measured by peptide mass fingerprinting of collagen // STAR Sci. Technol. Archaeol. – 2017. – Vol. 3. – P. 15-27.

25 Solazzo C., Wilson J., Dyer J.M. et al. Modeling deamidation in sheep α-keratin peptides and application to archeological wool textiles // Anal. Chem. – 2014. – Vol. 86. – P. 567-575.

26 Warinner C. et al. Direct evidence of milk consumption from ancient human dental calculus // Sci. Rep. – 2014. – Vol. 4. –P. 7104.

27 Cappellini E., Jensen L.J., Szklarczyk D., Ginolhac A., da Fonseca R.A.R., et al. Proteomic analysis of a Pleistocene mammoth femur reveals more than one hundred ancient bone proteins // J. Proteome Res. -2012. – Vol. 11. – P. 917–926

28 Demarchi B., Hall S., Roncal-Herrero T., et al. Protein sequences bound to mineral surfaces persist into deep time // Elife – 2016. – Vol. 5.

29 Le Meillour L., Zirah S., Zazzo A., et al. Palaeoproteomics gives new insight into early southern African pastoralism // Sci Rep. – 2020. – Vol. 10. –P 1-14.

30 Tyanova et al. The MaxQuant computational platform for mass-spectrometry-based shotgun proteomics // Nature Protocols. – 2016. – Vol. 11. – P. 2301-2319.

31 Coutu A.N., Taurozzi A.J., Mackie M., et al. Palaeoproteomics confirm earliest domesticated sheep in southern Africa ca. 2000 BP // Sci Rep. – 2021. – Vol. 11. – P. 6631.

32 Zeder M.A. Domestication and early agriculture in the Mediterranean Basin: Origins, diffusion, and impact // Proc. Natl. Acad. Sci. – 2008. – Vol. 105. – P. 11597-11604.

33 Olivieri C., Ermini L., Rizzi E., et al. Phylogenetic position of a copper age sheep (Ovis aries) mitochondrial DNA // PLoS One. – 2012. – Vol. 7(3). –P. e33792.

34 Hong C., Jiang H., Lü E. et al. Identification of milk component in ancient food residue by proteomics. PLoS One. – 2012. – Vol. 7. – P. 1-7.

35 Azémard C., Zazzob A., Mariea A., Lepetzb S., Debaine-Francfortc C., Idrissd SZ. Animal fibre use in the Keriya valley (Xinjiang, China) during the Bronze and Iron Ages: A proteomic approach // Journal of Archaeological Science. – 2019. – Vol. 110. – P. 104996.

36 Prendergast M.E., Janzen A., Buckley M., Grillo K.M. Sorting the sheep from the goats in the Pastoral Neolithic: morphological and biomolecular approaches at Luxmanda, Tanzania // Archaeol. Anthropol. Sci. – 2019. – Vol. 11. – P. 3047-3062.

37 Buckley M., Kansa S.W. Collagen fingerprinting of archaeological bone and teeth remains from Domuztepe, South Eastern Turkey // Archaeol. Anthropol. Sci. – 2011. – Vol. 3. – P. :271-280.

38 Le Meillour L., Zazzo A., Lesur J., et al. Identification of degraded bone and tooth splinters from arid environments using palaeoproteomics // Palaeogeogr Palaeoclimatol Palaeoecol. – 2018. – Vol. 511. – P. 472-482.

39 McKittrick J., Chen P.Y., Bodde S. G., Yang W., Novitskaya E.E., Meyers M.A. The Structure, Functions, and Mechanical Properties of Keratin // JOM Journal of the Minerals, Metals and Materials Society. – 2012. – Vol. 64. – P. 449–68.

40 Solazzo C., Wadsley M., Dyer J.M., Clerens S., Collins M.J., Plowman J. Characterisation of novel α-keratin peptide markers for species identification in keratinous tissues using mass spectrometry // Rapid Commun Mass Spectrom. – 2013. – Vol. 27. – P. 2685-2698.

41 Hollemeyer K., Altmeyer W., Heinzle E., Pitra, C. Matrix- assisted Laser Desorption/Ionization Time-of-flight Mass Spectrometry Combined with Multidimensional Scaling, Binary Hierarchical Cluster Tree and Selected Diagnostic Masses Improves Species Identification of Neolithic Keratin Sequences from Furs of the Tyrolean Iceman Oetzi // Rapid Communications in Mass Spectrometry. – 2012. – Vol. 26. – P. 1735–1745.

42 Charlton S., Ramsoe A., Collins M. et al. New insights into Neolithic milk consumption through proteomic analysis of dental calculus// Archaeol. Anthropol. Sci. – 2019. – Vol. 11. – P. 6183-6196.