Journal of Proteomics xxx (xxxx) 104187

Contents lists available at ScienceDirect

Journal of Proteomics

journal homepage: www.elsevier.com/locate/jprot

Ancient enamel peptides recovered from the South American Pleistocene species *Notiomastodon platensis* and *Myocastor cf. coypus*

Fabio C.S. Nogueira[a, b, 1], Leandro Xavier Neves[c, 1], Caroline Pessoa-Lima[d], Max Cardoso Langer[e], Gilberto B. Domont[a], Sergio Roberto Peres Line[f], Adriana Franco Paes Leme[c], Raquel Fernanda Gerlach[d, *]

[a] Proteomics Unit, Institute of Chemistry, Federal University of Rio de Janeiro (UFRJ), Rio de Janeiro, RJ, Brazil
[b] Laboratory of Proteomics/LADETEC, Federal University of Rio de Janeiro (UFRJ), RJ, Rio de Janeiro, Brazil
[c] Brazilian National Biosciences Laboratory (LNBio), Brazilian Center for Research in Energy and Materials (CNPEN), Campinas, SP, Brazil
[d] Department of Basic and Oral Biology, Faculty of Dentistry of Ribeirão Preto, University of São Paulo, FORP/USP, Brazil
[e] Laboratory of Paleontology, Department of Biology, FFCLRP, University of São Paulo, FFCLRP/USP, Ribeirão Preto, SP, Brazil
[f] Biosciences Department, Piracicaba Dental School, University of Campinas, FOP/UNICAMP, Piracicaba, SP, Brazil

ARTICLE INFO

Keywords:
Enamel
Peptides
Ancient
Tooth
Fossils
Notiomastodon platensis
Myocastor cf. coypus

ABSTRACT

We used two fossil teeth from South American Pleistocene mammals to obtain subsuperficial acid etching samples. We employed samples from the species *Notiomastodon platensis* and *Myocastor cf. coypus* for the enamel etchings. The controls included an extant rodent (rat). After the first etching was discarded, a second 20-s etching (*i.e.*, subsuperficial) was directly collected with a ZipTip and injected into an LTQ Orbitrap Velos for MS analysis. The peptides were identified with different software programs that used Peptide Spectrum Match (PSM) and *de novo* sequencing including similarity search strategies. Most of the peptides that were recovered from the enamel of the fossils belonged to enamel-specific proteins. To our knowledge, this is the first study that has described the recovery of enamel peptide molecules from extinct South American taxa, indicating that enamel peptide data from late Pleistocene fossils can be employed as an additional parameter for phylogenetic analysis, and that the sample can be obtained by a very conservative acid etching, with almost no damage to the fossils.

Significance: This study shows that it is possible to obtain information based on plenty of ancient peptides recovered from subsuperficial enamel of fossil teeth from South American Pleistocene. The quality of the data suggests that peptides are likely the best preserved biomolecules under certain harsh environmental conditions. The recovery procedure only lasted 20 s and was minimally destructive to the fossils. This opens a myriad of new possibilities for the study of the past.

1. Introduction

Many gaps exist in the knowledge about extinct species and the fossil record of extant species. Depending on the fossilization conditions, dental enamel is sometimes the only preserved tissue among vertebrate specimens. The most striking aspects of this tissue are probably hardness and density, which are the highest among vertebrate mineralized tissues [1,2]. Mature enamel is an acellular tissue, and it changes little after it is fully developed [3]. Whereas the dental enamel mineral content by weight is 96%, the bone mineral content by weight is less than 60% (depending on bone type). Therefore, enamel is the best extracellular matrix for biomolecular preservation for long periods [4–7]. However, to appreciate the full potential (and also the limitations) of using enamel as the source of proteins and peptides for fossil studies, some knowledge on enamel formation, structure, and differences as compared to other mineralized tissues is essential.

Fully mineralized “mature enamel” consists mainly of tightly packed hydroxyapatite crystals that occupy 89% of its volume. Mature enamel has lost almost all of its water and protein content (up to ~4% and < 1%, respectively, by weight) [8] and is the only mineralizing matrix that is laid down by epithelial cells (ameloblasts) among extant vertebrates (except for some fish scales). During the early stage of

* Corresponding author at: Av. do Café, S/N FORP/USP CEP: 14040-904 Monte Alegre, Ribeirão Preto, SP, Brazil.
E-mail address: rfgerlach@forp.usp.br (R.F. Gerlach).
[1] These authors contributed equally.

https://doi.org/10.1016/j.jprot.2021.104187
Received 25 April 2020; Received in revised form 13 March 2021; Accepted 17 March 2021
1874-3919/

enamel synthesis, ameloblasts secrete a distinct set of proteins that differ from the proteins that are normally expressed by cells from other mineralized tissues (bone, dentin, and cementum) [9]. The enamel organic matrix is made up of very specific proteins, 90% of which are amelogenins. The latter proteins are believed to have originated in the common vertebrate ancestor at least 450—500 million years ago [10,11] and are almost exclusively found in developing enamel.

In enamel, the mineral crystals are not deposited "onto" a proteinaceous scaffold, which is the case during dentinogenesis and bone and cementum formation. In dentin, bone, and cementum, the protein scaffold will last for many years, sometimes during the entire animal's life. This is the reason why connective tissue proteins are expected to be found during MS studies on fossil bones. On the other hand, enamel proteins are short-lived. The proteinaceous scaffold of the forming enamel matrix is believed to serve as a guide for crystal elongation during enamel formation (amelogenesis). In this process, enamel proteinases slowly degrade the most abundant proteins (amelogenin, ameloblastin, and enamelin), which are finally almost completely substituted with tightly packed hydroxyapatite crystals [12]. Enamel matrix mineralization is known as enamel maturation, and it is a very long process that happens before the tooth appears in the animal's oral cavity. During enamel maturation, the protein-rich matrix is degraded, protein products exit the enamel, the enamel water content is reduced, matrix acidification due to hydroxyapatite crystal precipitation is buffered, and the mineral content increases significantly (from around 30% to 96% by weight) [13]. In the case of thick enamel, the substitution process (maturation stage) can take a long time, and the acquisition of a very high mineral content (~95% by weight) depends on the enamel-forming cells (ameloblasts) [14].

In recent years, enamel has aroused great interest and inspired increasing confidence as a reliable repository of ancient proteins and peptides [1,2,6,7] because it preserves peptides and proteins and enables the use of non-destructive superficial enamel etching to obtain enamel samples for subsuperficial layer analysis. Subsuperficial samples are obtained after the outer enamel first layer is removed with acid, to generate samples that are not contaminated by peptides from outer sources [1,2,4,5]. In such studies, no connective tissue proteins are detected in the samples, which indicates the excellent quality of sample preparation. Interestingly, investigation of human deciduous subsuperficial enamel (10–30 μm below outer enamel) by Synchrotron light X-ray fluorescence has provided more consistent results regarding the microelement constitution [15], which suggests that a smaller amount of exogenous ions can reach that far (fewer ions, which are much smaller than organic molecules, are incorporated, which means that peptides and molecules this size (between 2 and 3 kDa) cannot diffuse that far, either). Indeed, in our earlier studies, we hardly found any protein that was not an enamel-specific protein in the mass spectral data [1,2,6,7]. Regarding fossil teeth, changes in the composition of the different dental enamel depths await investigation.

One very relevant aspect of enamel is still not adequately known outside the dental field: bacteria or other microorganisms cannot invade normal dental enamel even though they can easily invade bone and dentin in postmortem specimens. The reason for this difference is that the latter connective calcified tissues have microscopic pores [16], whereas dental enamel only contains nanopores, which are three orders of magnitude smaller than the pores found in bone or dentin. Thus, the very enamel structure can physically block the passage of any microorganism.

Very recently, the scientific community has welcomed inspiring work on the enamel proteome indicating that the proteome of enamel recovered from teeth as old as 1.77 million years can shed new light on hominin evolution [17]. Therefore, at this point, and bearing the amazing MS/MS analysis capability and sensitivity in mind, we thought that a relevant contribution to paleoproteomics would be to propose and to validate new simple and minimally invasive experimental protocols for proteomic enamel analysis in fossils.

In this study, we have investigated two fossil teeth from South American Pleistocene species. The enamel peptides were obtained by means of a subsuperficial 20-s acid etching. The peptides were identified by different software; Peptide Spectrum Match (PSM) and *de novo* sequencing including similarity search strategies were applied. We expect that this new and minimally invasive analysis method will help other researchers in the field of paleoproteomics to start using this amazing tissue and to employ the pipeline described here.

2. Materials and methods

2.1. Institutional abbreviations

LPRP/USP – Laboratório de Paleontologia de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto-SP, Brazil (Laboratory of Paleontology of Ribeirão Preto).

2.2. Fossils used in this study

LPRP/USP 0985 (Fig. 1) - The first sample belongs to a *Myocastor cf. coypus* lower incisor. *Myocastor coypus* is a medium-sized, caviomorph rodent with semi-aquatic habits; it is presently common in southern South America. The sample belongs to a large sample of *Myocastor* fossils collected by the LPRP team in the calcareous cave known as Gruta do Ioiô, in the Chapada Diamantina area, in Bahia, Brazil [18,19]. Several faunal components of the Gruta do Ioiô have been dated, and the ages ranged from 30.3 to 8.3 ky BP [15,16], which encompasses the "upper" Pleistocene and the Greenlandian (Holocene). In particular, a *Myocastor cf. coypus* tooth, but not the one sampled here, has been dated as 19.9–20.2 ky BP [19].

LPRP/USP 0986 (Fig. 2) - The second sample analyzed herein consisted of a gomphothere tooth fragment with a bucco-lingual width of ~6 cm and enamel width of approximately 0.8 cm (Fig. 2). On the basis of recent revisions of South American mastodons provided by Mothé et al., the sample can be assigned to *Notiomastodon platensis*, [20,21]. *Notiomastodon platensis* is an extinct, large-sized proboscidean that was widespread in South America from the Pleistocene to the Holocene. The specimen was donated to Dr. Sergio R P Line by Guinea Brasil Camargo Cardoso between 2005 and 2008 and is now housed at LPRP. Information from the donor indicates that the material is likely to have come from Lagoa de Dentro, Puxinanã, PB, Brazil, but no further information has been provided. Lagoa de Dentro corresponds to a depression filled with water during the rainy season, which hosts fossiliferous materials with outstanding degree of preservation [22]. Kinoshita et al. [23] dated two gomphothere teeth from the area, to recover a mean age of 11.6 ky BP (*i.e.*, Pleistocene/Holocene boundary). Yet, because

Fig. 1. *Myocastor cf. coypus* fossilized incisor tooth. Lateral view. Bar = 1 cm.

Fig. 2. *Notiomastodon platensis* fossil molar fragment. Bar = 1 cm.

these kinds of deposits encompass a large time averaging, this is no more than a rough age estimate for the material.

2.3. Enamel etching

All the reagents were of analytical grade. The solvents that were used for nano-Liquid Chromatography tandem mass spectrometry (nanoLC-MS/MS) were MS grade, and their precise catalog numbers were: Formic acid (F0507) and Trifluoroacetic acid (TFA)(T6508), from Sigma Aldrich (St Louis, MO, USA); Methanol (LC-MS ultra chromosolv 14,262), from FLuka Analytical (Buchs, Switzerland); and Acetonitrile (LC-MS ultra chromosolv 34,967), from FLuka Analytical. Hydrochloric acid (HCl) was bidistilled in the Chemistry Department in our campus. The C18 resin-loaded Ziptip that was employed in this study came ready from the supplier (ZTC18S096; Millipore, Merck, Sigma Aldrich) and was used in a 0.5- to 10-μL adjustable pipette. Enamel was rinsed with ultrapure water before its surface was etched with 10% (vol/vol) HCl for 20 s. A tape (Magic Tape, 3 M Corporation, Campinas, SP, Brazil) with central perforation was placed on enamel, to limit spread of the acid that was used for the etching. Fig. 3 shows one of the teeth with the tape in place. Approximately 10 μL of 10% HCl (vol/vol) was placed on enamel, within the limits of the perforation. This acid was the "first biopsy", and was discarded. A Ziptip was then quickly conditioned (with eight up and down movements with methanol) and immediately used to make up and down movements with 10% HCl on top of the tooth for as long as 20 s (normally eight up and down movements with the liquid). Next, the volume within the Ziptip was discarded, and the Ziptip was washed with 0.1% (vol/vol) TFA eight times; all the washes were discarded. After that, 10 μL of 80% (vol/vol) acetonitrile containing 0.1% formic acid was used for the elution. The peptides that were bound to the resin were eluted during this step, and this peptide mixture was left inside a Class 100 hood

Fig. 3. *Myocastor cf. coypus* fossilized incisor tooth photographed from the superficial (buccal) side, showing the plastic tape that was used to limit the acid attack on the structure. Bar = 1 cm.

that was "on" until acetonitrile had evaporated completely. The tubes were then closed, parafilm was used to wrap them, and the peptides remained lyophilized until analysis. A procedure flow chart is presented in Fig. 4, and a video is also available (videolink). Two etchings were conducted in this way. The first etching, designated "superficial" etching, was not analyzed in any of the samples. The second etching, designated "subsuperficial" etching, was analyzed for the two fossil samples (mastodons tooth – sample A2; *Myocastor cf. coypus* tooth – sample A5) and the positive control, which was a rat (*Rattus norvegicus* – sample A4) upper incisor tooth that was processed exactly as the fossil samples. This method closely resembled the method that was employed earlier [6,7], but a shorter etching time was applied (20 s), and the peptides were directly collected from enamel and not from a solution obtained by etching. In short, the peptides were allowed to bind to the resin by mixing the acid solution with up and down movements that were directly made on enamel with the aid of the pipette through the Ziptip. With each up and down movement, the peptides were exposed to the resin.

After the two enamel etchings, the pictures shown in **Figs. 5** and **6** were taken; the area that was demineralized by the acid is clearly indicated in these two pictures. In the *Myocastor cf. coypus* tooth (Fig. 5), the tape window is clearly seen as a white area after the etching, indicating that the entire superficial layer was removed. This superficial layer has an orange aspect due to the iron deposit that is characteristic of rodent teeth [24]. In rats, this superficial layer is present in the outermost 5 μm. In the *Myocastor cf. coypus* tooth, this may also be true, but so far we have not measured the depths of the removed layers or determined the iron content of the surface of this rodent fossil (such studies are underway). (See Fig. 7.)

No previous digestion with trypsin or any other treatment was applied to the samples. No proteinase or phosphatase inhibitors were used in the etching solution. All the samples were analyzed twice.

2.4. Nano-LC MS/MS analysis

The peptide mixtures were dissolved in 10 μL of 0.1% formic acid, and 4 μL of the resulting solution was loaded onto a pre-column (2-cm length, 200-μm inner diameter, packed in-house with ReproSil-Pur C18-AQ 5-μm resin), which was followed by fractionation on a New Objective PicoFrit analytical column (25-cm length, 75-μm inner diameter, packed in-house with ReproSil-Pur C18-AQ 3-μm resin). The samples were injected in a nano-LC–MS/MS system consisting of an EASY II-nano LC system (Proxeon Biosystem) coupled to a nanoESI LTQ-Orbitrap Velos mass spectrometer (Thermo Fisher Scientific). The peptides were eluted by using a gradient from 95% phase A (0.1% formic acid) to 40% phase B (0.1% formic acid, 95% acetonitrile) for 47 min, 40–100% phase B for 8 min, and 100% B for 5 min (total of 60 min at a flow rate of 250 nL/min). After each run, the column was washed with 90% phase B and re-equilibrated with phase A. The m/z spectra were acquired in the positive mode; survey MS scan and tandem mass spectra (MS/MS) acquisition were applied. An MS scan (350–2000 m/z) in the Orbitrap mass analyzer set at a resolution of 60,000 at 400 m/z, 1×10^6 automatic gain control target, and 500-ms maximum ion injection time was followed by MS/MS of the ten most intense multiply charged ions by HCD at 30,000 signal threshold, 5000 automatic gain control target, 300-ms maximum ion injection time, 2.0 m/z isolation width, 30 normalized collision energy, and dynamic exclusion enabled for 45 s with a repeat count of 1.

2.5. Data analysis and database search (Proteome Discoverer 2.1)

The raw data were inspected in Xcalibur v.2.1 (Thermo Scientific). Database searches were performed by using the Proteome Discoverer with Sequest™ algorithm against entries from Elephantidae (27,516 entries) and Rodentia (471,755) taxonomy downloaded from the

Fig. 4. Peptide recovery protocol Flow Chart.

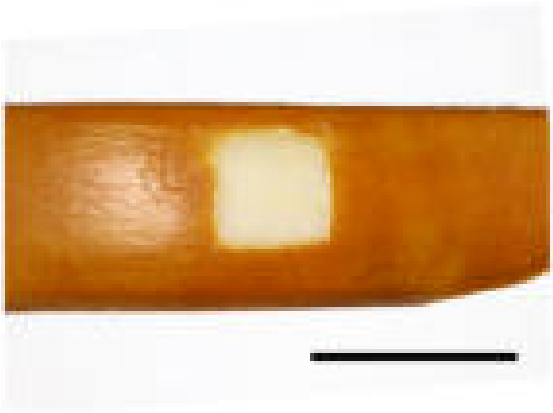

Fig. 5. *Myocastor cf. coypus* fossilized incisor tooth photographed from the superficial (buccal) side at a higher magnification, after removal of the tape that limited the spread of the acid. Note that the superficial layer containing iron was removed, and the enamel etched area is now shown in white. Bar = 0.5 cm.

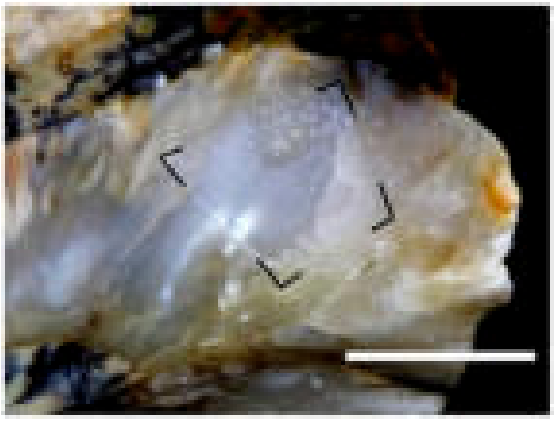

Fig. 6. *Notiomastodon platensis* fossilized molar tooth after removal of the tape that had been put in place to limit the etching area (black lines indicate the corners).

UniProt database in February 2020. Rodentia was chosen because *Myocastor* belongs to this order. Elephantidae was selected because it is the living group that is phylogenetically the closest to *Notiomastodon*. Elephantidae has protein sequences of species such as the African and the Indian elephant, whereas Rodentia has protein sequences of mouse, rat,

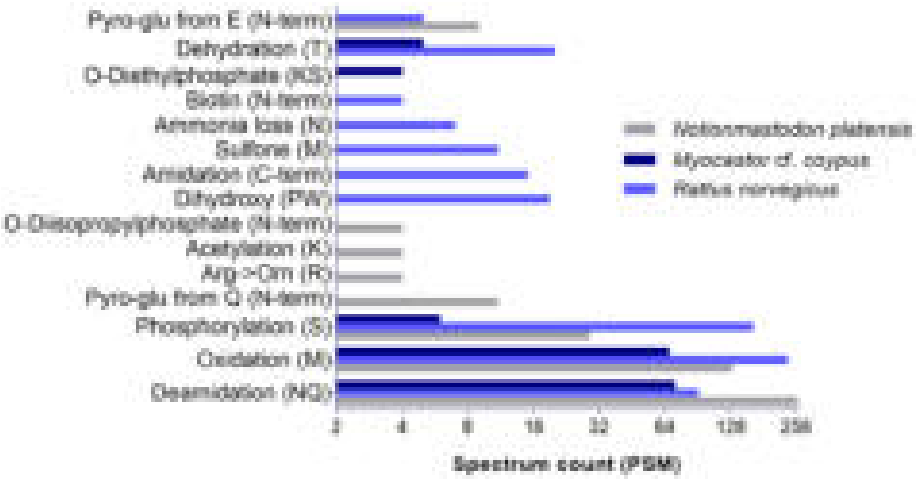

Fig. 7. Peptide modifications identified by using PEAKS PTMs. The number of spectra assigned to peptides with chemical modifications in the *Rattus norvegicus* (A4), *M. cf. coypus* (A5), and *N. platensis* (A1) samples.

golden hamster, and guinea pig, for example. The searches were performed with the following parameters: precursor accuracy of 10 ppm, fragment accuracy of 0.05 Da, and variable modification of oxidized methionine and deamidation of glutamine and asparagine. The number of proteins and protein groups and the number of peptides were estimated by using the Proteome Discoverer cutoff limits, maximum parsimony, false discovery rates <1%, and peptide rank 1.

2.6. De novo sequencing and similarity search (Rapid Novor and PepExplorer from PatternLab for proteomics)

The raw data were converted to mgf by using the MS Converter tool from the ProteoWizard platform [25]. The mgf files were submitted to automatic *de novo* sequencing by employing the Rapid Novor software (www.rapidnovor.com) with the parameters MS accuracy 10 ppm, MS/MS accuracy 0.05 Da for HCD, and variable modification of oxidized methionine. The output files were loaded on the PepExplorer module from PatternLab for Proteomics software. The target-decoy databases were generated by using PatternLab to include a reversed version of each sequence that was found in the previously described databases plus the versions from 127 common mass spectrometry contaminants. The searches were performed with the following parameters: sequences formed by at least six amino acids, minimal identity of 0.75, denovo score cutoff equal or superior to 80, FDR < 1%, and maximum parsimony checked.

2.7. De novo sequencing-assisted database search by using PEAKS

The raw files were analyzed by using PEAKS Studio X (Bioinformatics Solutions Inc., Canada) [26]. The *de novo* sequencing was first performed without restriction to cleavage specificity; methionine oxidation was used as variable modification, and the mass error tolerances were set to 10 ppm and 0.02 Da for precursors and product ions, respectively. Next, the PEAKS search engine interrogated canonical and isoform protein sequences of *Rattus norvegicus* (30,198 sequences) Rodentia (479,884 sequences) and Elephantidae (27,507 sequences) downloaded from UniProt (12/03/2020). PEAKS PTM [27] and SPIDER [28] were activated for post-translational modification (PTM) profiling and sequence variation analysis (minimal relative intensity of position-determining fragment ions in the MS/MS spectra was set to > 1%). The database search results were filtered to keep false discovery rate (FDR) ≤ 0.01 for peptides and proteins and at least one significant unique peptide per protein group. The *de novo* sequences with average local confidence (ALC) ≥98% and length ≥ 7 were submitted to BLASTp against the NCBInr *Rattus norvegicus* (taxid: 10116), Rodentia (taxid: 9989), and Elephantidae (taxid: 9780) databases.

3. Results & discussion

3.1. MS-based identification of enamel proteins in fossils

The MS-based *de novo* sequencing and similarity search conducted with Rapid Novor and PepExplorer from Pattern Lab software, respectively, identified the proteins that are presented in **Supplemental Material** Tables 1, **2**, and **3**. In sample A1 (*Notiomastodon platensis*), we identified the proteins Enamelin, COL17A1, AHSG, NEDD1, Amelogenin X, and MMP20. In sample A4 (*Rattus norvegicus*), we identified Ameloblastin, FAM171A1, Amelogenin X, and Actb with a higher score. In sample A5 (*Myocastor cf. coypus*), we identified Amelogenin, Enamelin, and STK11IP. We identified two BSA peptides in sample A1 (Observation: we calibrated the Instrument with BSA and performed washout runs, but they might not have been enough).

Data analysis in the PEAKS Studio environment relies on a first stage of automated *de novo* sequencing of peptide spectra. The data that were filtered by the default *de novo* score threshold ≥50 resulted in 615, 665, and 1835 peptides being identified in samples A4 (*Rattus norvegicus*), A5 (*Myocastor cf. coypus*), and A1 (*Notiomastodon platensis*), respectively. We then searched the peptide sequences from the *de novo* sequencing in the protein databases for confident assignment (FDR ≤ 1%) to proteins of *Rattus norvegicus* (A4, 236 PSM to 138 pep-

Table 1
Performance of PEAKS Studio for identification of peptides from fossil samples.

Sample	Organism	UniProt taxonomy searched	*de novo* peptides	Database search: Peptides (PSM)	PTM & mutation search: Peptides (PSM)	*De novo* only peptides
A1	*Notiomastodon platensis*	Elephantidae	1835	60 (110)	258 (509)	82
A4	*Rattus norvegicus*	Rat	615	138 (236)	394 (639)	1
A5	*Myocastor cf. coypus*	Rodentia	665	30 (73)	136 (238)	24

tides), *Rodentia* (A5, 73 PSM to 30 peptides), and *Elephantidae* (A1, 110 PSM to 60 peptides) taxon (Table 1). All the results that we obtained by using PEAKS for samples A1 (*N. platensis*), A4 (rat), and A5 (*M. cf. coypus*) are presented in **Supplemental Material Tables 4**, **5**, and **6**, respectively.

We found dozens of peptides that are specific to enamel proteins. Regarding amelogenins, Stewart et al. [6] already described that mostly peptides corresponding to the N- and C-terminus were detected, and that not many peptides corresponded to the middle part of the protein (alignment not shown).

Next, we enabled the PEAKS PTM and SPIDER algorithms to maximize identification of chemically modified peptides (up to 313) and homologous sequences by using the spectral data that had not been matched in the previous database search. This approach resulted in a 2.8- to 4.5-fold increase in the number of detected peptides (Table 1). Fig. 6 illustrates that, in addition to common methionine oxidation, deamidation of asparagine and glutamine residues was the most frequent modification in the fossil samples. We detected serine phosphorylation in all of the samples, but it was more prominent in rat. Moreover, we detected O-diisopropylphosphate, pyroglutamic acid (from Q), lysine acetylation, and arginine to ornithine conversion exclusively in *N. platensis* proteins. We detected O-diethylphosphate (K/S) uniquely in *M. cf. coypus*, whereas we detected biotin (N-term), ammonia loss, methionine sulfone, C-terminal amidation, dihydroxy-proline, and tryptophan uniquely in rat proteins. In addition, we detected pyroglutamic acid (from E) in the *N. platensis* and rat samples, whilst we observed threonine dehydration in both rat and *M. cf. coypus*.

Potential contaminant peptides found are the following: in sample A1: BSA (two peptides), A4: BSA (five peptides); A5: BSA (four peptides), and streptavidin (one peptide).

Notably, a group of PTMs detected exclusively in fossil samples might indicate protein degradation of ancient samples. For instance, arginine to ornithine conversion in early Pleistocene enamel proteins has been reported in another study [29]. Our data suggest that the presence of O-diethyl/O-diisopropylphosphate, which we exclusively detected in the *M. coypus* and *N. platensis* samples, respectively, might have similar meaning. Although pyroglutamic acid (pGlu) formation from glutamine or glutamic acid occurs *in vivo*, these modifications can probably accumulate in fossil samples because they can also take place spontaneously. In fact, we detected pGlu from Q in *N. platensis*, whereas pGly from E was present in both the *N. platensis* and rat samples, but with lower frequency in the latter sample. Similarly, when we look at PTM frequency, deamidation comprised 59% and 46% of all the modifications that we detected in *N. platensis* and *M. coypus,* respectively, whilst they accounted for 16% of the modifications in the rat sample, thereby corroborating previous work reporting increased deamidation levels in archaeological samples [30]. On the other hand, PTM exclusive to the rat sample is more likely to be related to *in vivo* modification.

We assigned all the phosphopeptides that we detected in the fossil samples to amelogenin, whereas we detected amelogenin- and ameloblastin-derived phosphopeptides in modern rat samples. Notably, all the phosphosites that we identified in these sequences were constrained at serine residues (Fig. 8). Ser-32 phosphorylation is conserved in human, bovine, porcine, rat, and mouse amelogenin (among others), so our data revealed that this phosphorylation site was preserved in the fossils of South American extinct species, under the particular conditions found in this continent. Additionally, we detected two phosphorylation sites in the rat ameloblastin, Ser-48 and Ser-241. Welker et al. [17] found serine phosphorylation, including the same amelogenin and ameloblastin phosphosites detected in our samples, in homo samples that are estimated to be 1.77 million years old. Thus, fossil record shows that posttranslational modifications are maintained under several conditions. In our study, we mention Ser-32, which is the same Serine residue known as Ser-17 and in some species Ser-16 in earlier protein studies [31].

Next, we explored the sequence features of amelogenin that we identified in our samples by aligning them with the reference human sequences. Most importantly, because the databases that we used had multiple entries for amelogenin, we assigned the rat, *M. cf. coypus*, and *N. platensis* peptide hits to more than one sequence, which resulted in the identification of two, five, and two protein groups in these samples, respectively (Fig. 9). Protein grouping derives the minimal protein list that is sufficient to account for the observed peptide identifications. Briefly, if two or more protein sequences match the same set of peptides, they should be reported as a group because the information at the peptide level is not enough to differentiate between them. Therefore, this required that each protein group that we identified had at least one unique peptide to be valid.

Human amelogenin is encoded in sex chromosomes (AMELX and AMELY), so that sexual dimorphism (Fig. 9, highlighted in red) is translated into proteoforms. We observed that most of the multiple protein group matches for *Rattus norvegicus* and *M. cf. coypus* were related to sequence variation, including alternative splicing. In both rat (A4) and *M. cf. coypus* (A5), our findings confirmed the co-occurrence of the "SYEVLTP" and "SYEKSHSXAINXDRTALVLTP" sequences, which resulted from exon 3 skipping and insertion, respectively (Fig. 8). Although we did not achieve the same coverage for exon 3-encoded amino acids in the *N. platensis* sample (A1), identification of the "SYEVLTPLK" peptide (*via de novo* sequencing) simultaneously with "AGVLTPLK" might indicate that alternative spliced proteoforms occur (Fig. 9). In addition to alternative splicing, sequence variation in rat (A4) amelogenin included simultaneous detection of peptides containing either Tyr or Ala in position 64, and Pro or Phe in position 65 (using UniProt P632278 as reference). In *M. coypus* (A5), sequence variation included the co-occurrence of Leu and Phe in position 53, and Tyr or Pro in position 64 (UniProt H0W138_CAVPO as reference). In *N. platensis* (A1), we observed a single variation at the protein N-terminal as indicated by simultaneous identification of Pro or Lys residues in position 18 (UniProt G3U0C2_LOXAF as reference). It is noteworthy that the experimental sequence variations highlighted in Fig. 8 (red boxes) encompasses only the instances where two distinct sequences covering the same protein backbone region were identified as co-occurring in the same sample. The degree of conservation or variation among species is indicated by the alignment symbols below the sequences.

The quality of the data suggests that peptides are likely the best-preserved biomolecules under certain harsh environmental conditions. The recovery procedure lasted only 20 s and was minimally destructive

Fig. 8. Sequence logo of phosphorylation sites in amelogenin and ameloblastin. Phosphopeptides identified *via* PEAKS PTM. The frequency of each residue in peptides is proportional to the number of bits (y axis). The phosphorylation site is 0 on the x axis flanked by 3 amino acid residues.

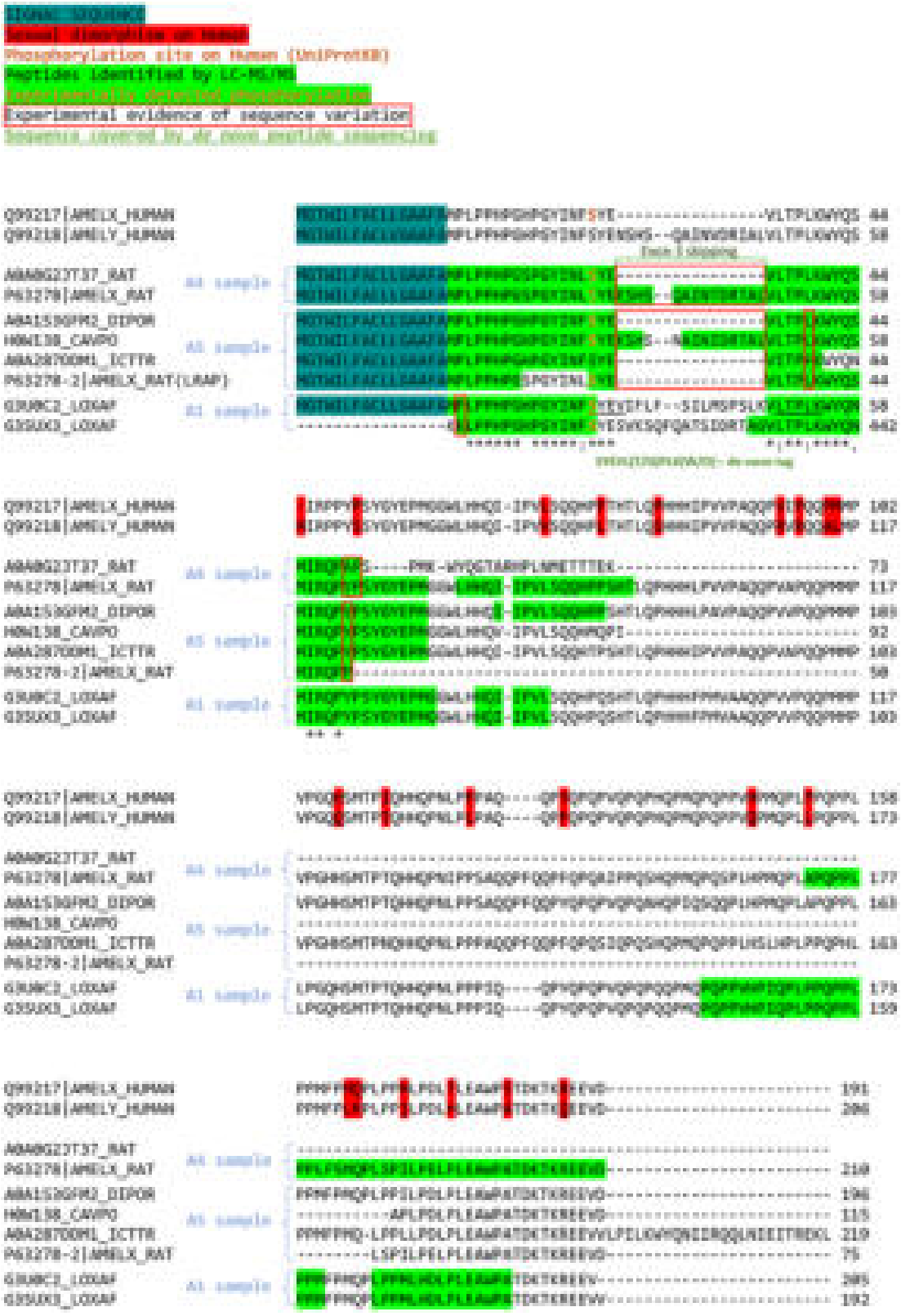

Fig. 9. Alignment of the sequence of amelogenin identified from our data with the reference human sequences. The colors used for identification are explained in the upper left side of the figure.

to the fossils. We hope that the precise description of the 20-s enamel etching and the quality of the data will enable other researchers to use this extremely conservative way to obtain biological molecules, especially in samples that are unlikely to have ancient DNA preserved.

Among the proteins that we found in our study, we did not detect peptides of the major collagen types, particularly collagen type I, which is the major structural protein of mineralized connective tissues. This agreed with the different constitution of enamel. The fact that enamel contains a different set of proteins is a well-known fact, and

this fundamental observation was made with the ground-setting studies conducted by a group of biochemists that studied in depth immature and mature enamel matrixes [32,33]. Furthermore, when fossil enamel and fossil dentin were analyzed with appropriate care to avoid contamination, no amino acid residues typical of collagen type I were found in the fossilized enamel [34]. Regarding the finding of collagen XVII, interestingly we found four peptides for this protein (please see Supplemental Material Table 1), and this protein has indeed been described to be essential for normal enamel formation [35].

Determining the phylogenetic relationship of fossil specimens is one of the major goals of paleontology. The relatively recent improvement in the protocols for the recovery of DNA and protein fragments (mainly collagen I) from fossil material has sparked the interest of researchers in this area [36–38]. DNA and protein not only confirm previous analysis based on bone and tooth morphology, but also have the potential to increase the precision of phylogenetic relationships. As mentioned earlier, the preservation of ancient proteins and DNA recovered from fossilized bone is limited by the fact that this tissue contains a net of microscopic canaliculi left by the osteocyte prolongments, which can enhance aerobic decomposition and allow bacterial invasion with consequent organic molecule degradation. The high mineral content and denser structure of enamel protect its organic content from degradation. Besides phylogenetic relationships, enamel proteome analysis can reveal the sex of specimens [6], opening the exciting possibility for the analysis of sexually dimorphic characteristics in fossil specimens.

4. Conclusions

Enamel-specific peptides were recovered from South American Pleistocene mammals. The recovery procedure was extremely simple, lasted only 20 s, and consisted of diluted acid etching followed by direct peptide collection from enamel with Ziptip. The method yielded high-quality samples, which allowed dozens of peptides from the enamel proteins to be identified, while contamination by unexpected peptides was minimal. Therefore, ancient biomolecules that have been stored in enamel for thousands to millions of years can be obtained with minimal damage to the fossils.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jprot.2021.104187.

CRediT authorship contribution statement

Fabio C.S. Nogueira: Software, Formal analysis, Validation, Data curation, Writing - original draft. **Leandro Xavier Neves:** Validation, Formal analysis, Visualization, Writing - original draft. **Caroline Pessoa-Lima:** Conceptualization, Methodology, Investigation, Resources, Visualization, Writing - original draft. **Max Cardoso Langer:** Data curation, Writing - original draft, Funding acquisition. **Gilberto B. Domont:** Resources, Data curation. **Sergio Roberto Peres Line:** Conceptualization, Data curation, Writing - review & editing, Visualization. **Adriana Franco Paes Leme:** Methodology, Formal analysis, Writing - review & editing. **Raquel Fernanda Gerlach:** Conceptualization, Funding acquisition, Writing - original draft, Supervision.

Acknowledgments

This study was supported by FAPESP (The State of São Paulo Research Foundation) and CNPq (The Brazilian Council for Research Development).

References

[1] I.M. Porto, H.J. Laure, R.H. Tykot, F.B. De Sousa, J.C. Rosa, R.F. Gerlach, Recovery and identification of mature enamel proteins in ancient teeth, 2021, https://doi.org/10.1111/j.1600-0722.2011.00885.x.

[2] I.M. Porto, H.J. Laure, F.B. de Sousa, J.C. Rosa, R.F. Gerlach, New techniques for the recovery of small amounts of mature enamel proteins, J. Archaeol. Sci. 38 (2011) 3596–3604, https://doi.org/10.1016/j.jas.2011.08.030.

[3] S.I. Djomehri, S. Candell, T. Case, A. Browning, G.W. Marshall, W. Yun, S. H. Lau, S. Webb, S.P. Ho, Mineral density volume gradients in Normal and diseased human tissues, PLoS One 10 (2015) e0121611, https://doi.org/10.1371/journal.pone.0121611.

[4] S.R. Peres Line, L.P. Bergqvist, Enamel structure of paleocene mammals of the São José de Itaboraí basin, Brazil. "Condylarthra" Litopterna, Notoungulata, Xenungulata and Astrapotheria, J. Vertebr. Paleontol. 25 (2005) 924–928, https://doi.org/10.1671/0272-4634(2005)025[0924:ESOPMO]2.0.CO;2.

[5] L.L. Ramenzoni, S.R.P. Line, Automated biometrics-based personal identification of the Hunter-Schreger bands of dental enamel, Proc. R. Soc. B Biol. Sci. 273 (2006) 1155–1158, https://doi.org/10.1098/rspb.2005.3409.

[6] N. Stewart, R. Fernanda, R.L. Gowland, K.J. Gron, J. Montgomery, Sex determination of human remains from peptides in tooth enamel, 114, 2017, pp. 13649–13654, https://doi.org/10.1073/pnas.1714926115.

[7] N.A. Stewart, G.F. Molina, J.P. Mardegan Issa, N.A. Yates, M. Sosovicka, A. R. Vieira, S.R.P. Line, J. Montgomery, R.F. Gerlach, The identification of peptides by nanoLC-MS/MS from human surface tooth enamel following a simple acid etch extraction, RSC Adv. 6 (2016) 61673–61679, https://doi.org/10.1039/c6ra05120k.

[8] C.E. Smith, Cellular and chemical events during enamel maturation, Crit. Rev. Oral Biol. Med. 9 (1998) 128–161.

[9] D. Deutsch, J. Catalano-Sherman, L. Dafni, S. David, A. Palmon, Enamel matrix proteins and ameloblast biology, Connect. Tissue Res. 32 (1995) 97–107, https://doi.org/10.3109/03008209509013710.

[10] J.-Y. Sire, S. Delgado, D. Fromentin, M. Girondot, Amelogenin: lessons from evolution, Arch. Oral Biol. 50 (2005) 205–212, https://doi.org/10.1016/j.archoralbio.2004.09.004.

[11] S. Delgado, D. Casane, L. Bonnaud, M. Laurin, J. Sire, M. Girondot, Molecular Evidence for Precambrian Origin of Amelogenin , the Major Protein of Vertebrate Enamel, 1998, pp. 2146–2153.

[12] A. Nanci, Ten Cate'S Oral Histology: Development, Structure & Function, 7th ed., 2008 https://books.google.com.br/books/about/Ten_Cate_S_Oral_Histology_Development_St.html?id=qOrXMgEACAAJ&redir_esc=y. (Accessed 21 August 2020).

[13] J.P. Simmer, J.C.C. Hu, Expression, structure, and function of enamel proteinases, Connect. Tissue Res. 43 (2002) 441–449, https://doi.org/10.1080/03008200290001159.

[14] I.M. Porto, J. Merzel, F.B. de Sousa, L. Bachmann, J.A. Cury, S.R.P. Line, R. F. Gerlach, Enamel mineralization in the absence of maturation stage ameloblasts, Arch. Oral Biol. 54 (2009) 313–321, https://doi.org/10.1016/j.archoralbio.2009.01.007.

[15] C. de Souza-Guerra, R.C. Barroso, A.P. de Almeida, I.T.A. Peixoto, S. Moreira, F.B. De Sousa, R.F. Gerlach, Anatomical variations in primary teeth microelements with known differences in lead content by micro-synchrotron radiation X-ray fluorescence (μ-SRXRF) - a preliminary study, J. Trace Elem. Med. Biol. 28 (2014) 186–193, https://doi.org/10.1016/j.jtemb.2014.01.007.

[16] I. Mjor, O. Jeferskov, Human Oral Embryology and Histology, First ed., Blackwell Science Ltd, 1986.

[17] F. Welker, J. Ramos-Madrigal, P. Gutenbrunner, M. Mackie, S. Tiwary, R. Rakownikow Jersie-Christensen, C. Chiva, M.R. Dickinson, M. Kuhlwilm, M. de Manuel, P. Gelabert, M. Martinón-Torres, A. Margvelashvili, J.L. Arsuaga, E. Carbonell, T. Marques-Bonet, K. Penkman, E. Sabidó, J. Cox, J. V. Olsen, D. Lordkipanidze, F. Racimo, C. Lalueza-Fox, J.M. Bermúdez de Castro, E. Willerslev, E. Cappellini, The dental proteome of Homo antecessor, Nature. 580 (2020) 235–238, https://doi.org/10.1038/s41586-020-2153-8.

[18] M.C. Castro, F.C. Montefeltro, M.C. Langer, The quaternary vertebrate fauna of the limestone cave Gruta do Ioiô, northeastern Brazil, Quat. Int. 352 (2014) 164–175, https://doi.org/10.1016/j.quaint.2014.06.038.

[19] E. Eltink, M. Castro, F.C. Montefeltro, M.A.T. Dantas, C.S. Scherer, P.V. de Oliveira, M.C. Langer, Mammalian fossils from Gruta do Ioiô cave and past of the Chapada Diamantina, northeastern Brazil, using taphonomy, radiocarbon dating and paleoecology, J. S. Am. Earth Sci. 98 (2020) 102379, https://doi.org/10.1016/j.jsames.2019.102379.

[20] D. Mothé, L.S. Avilla, M. Cozzuol, G.R. Winck, Taxonomic revision of the quaternary gomphotheres (Mammalia: Proboscidea: Gomphotheriidae) from the South American lowlands, Quat. Int. 276–277 (2012) 2–7, https://doi.org/10.1016/j.quaint.2011.05.018.

[21] D. Mothé, M.P. Ferretti, L.S. Avilla, Running over the same old ground: stegomastodon never roamed South America, J. Mamm. Evol. 26 (2019) 165–177, https://doi.org/10.1007/s10914-017-9392-y.

[22] J. Almeida, J. Macario, R. Schultz, J. Queiroz, Lagoa de Dentro: um jazigo de Megafauna Pleistocênica em Puxinanã - Paraíba, Rev. Bras. Paleontol. (2001) 78–79.

[23] A. Kinoshita, A. Magnólia Franca, J.A.C. De Almeida, A.M. Figueiredo, P. Nicolucci, C.F. Carlos, O. Baffa, ESR dating at K and X band of northeastern Brazilian megafauna, Appl. Radiat. Isot. (2005) 225–229 Elsevier Ltd, https://doi.org/10.1016/j.apradiso.2004.08.007.

[24] R.A. Saiani, I.M. Porto, E. Marcantonio Junior, J.A. Cury, F.B. de Sousa, R. F. Gerlach, Morphological characterization of rat incisor fluorotic lesions, Arch. Oral Biol. 54 (2009) 1008–1015, https://doi.org/10.1016/j.

archoralbio.2009.08.009.
[25] M.C. Chambers, B. MacLean, R. Burke, D. Amodei, D.L. Ruderman, S. Neumann, L. Gatto, B. Fischer, B. Pratt, J. Egertson, K. Hoff, D. Kessner, N. Tasman, N. Shulman, B. Frewen, T.A. Baker, M.Y. Brusniak, C. Paulse, D. Creasy, L. Flashner, K. Kani, C. Moulding, S.L. Seymour, L.M. Nuwaysir, B. Lefebvre, F. Kuhlmann, J. Roark, P. Rainer, S. Detlev, T. Hemenway, A. Huhmer, J. Langridge, B. Connolly, T. Chadick, K. Holly, J. Eckels, E.W. Deutsch, R.L. Moritz, J.E. Katz, D.B. Agus, M. MacCoss, D.L. Tabb, P. Mallick, A cross-platform toolkit for mass spectrometry and proteomics, Nat. Biotechnol. 30 (2012) 918–920, https://doi.org/10.1038/nbt.2377.
[26] J. Zhang, L. Xin, B. Shan, W. Chen, M. Xie, D. Yuen, W. Zhang, Z. Zhang, G. A. Lajoie, B. Ma, PEAKS DB: De novo sequencing assisted database search for sensitive and accurate peptide identification, Mol. Cell. Proteomics 11 (2012), https://doi.org/10.1074/mcp.M111.010587.
[27] X. Han, L. He, L. Xin, B. Shan, B. Ma, PeaksPTM: mass spectrometry-based identification of peptides with unspecified modifications, J. Proteome Res. 10 (2011) 2930–2936, https://doi.org/10.1021/pr200153k.
[28] Y. Han, B. Ma, K. Zhang, Spider: software for protein identification from sequence tags with de novo sequencing error, J. Bioinforma. Comput. Biol. 3 (2005) 697–716, https://doi.org/10.1142/S0219720005001247.
[29] E. Cappellini, F. Welker, L. Pandolfi, J. Ramos-madrigal, P.L. Rüther, A.K. Fotakis, D. Lyon, J.V. Moreno-mayar, M. Bukhsianidze, R.R. Jersie-christensen, M. Mackie, R. Ferring, M. Tappen, E. Palkopoulou, M.R. Dickinson, Early Pleistocene enamel proteome sequences from Dmanisi resolve Stephanorhinus phylogeny, Nature. 574 (2020) 103–107, https://doi.org/10.6084/m9.figshare.7212746.
[30] N.L. Van Doorn, J. Wilson, H. Hollund, M. Soressi, M.J. Collins, Site-specific deamidation of glutamine: a new marker of bone collagen deterioration, Rapid Commun. Mass Spectrom. 26 (2012) 2319–2327, https://doi.org/10.1002/rcm.6351.
[31] T. Takagi, M. Suzuki, T. Baba, K. Minegishi, S. Sasaki, Complete amino acid sequence of amelogenin in developing bovine enamel, Biochem. Biophys. Res. Commun. 121 (1984) 592–597, https://doi.org/10.1016/0006-291X(84)90223-7.
[32] J.D. Termine, D.A. Torchia, K.M. Conn, Enamel matrix: structural proteins, J. Dent. Res. 58 (1979) 773–781, https://doi.org/10.1177/00220345790580022901.
[33] J.D. Termine, A.B. Belcourt, P.J. Christner, K.M. Conn, M.U. Nylen, Properties of dissociatively extracted fetal tooth matrix proteins. I. Principal molecular species in developing bovine enamel, J. Biol. Chem. 255 (1980) 9760–9768 https://pubmed.ncbi.nlm.nih.gov/7430099/. (Accessed 24 August 2020).
[34] M.J. Glimcher, L. Cohen-solal, D. Kossiva, A. De Ricqles, S. Paleobiology, N. Spring, M.J. Glimcher, L. Cohen-solal, D. Kossiva, A. De Ricqles, Biochemical Analyses of Fossil Enamel and Dentin, 16, 1990, pp. 219–232.
[35] T. Asaka, M. Akiyama, T. Domon, W. Nishie, K. Natsuga, Y. Fujita, R. Abe, Y. Kitagawa, H. Shimizu, Type XVII collagen is a key player in tooth enamel formation, Am. J. Pathol. 174 (2009) 91–100, https://doi.org/10.2353/ajpath.2009.080573.
[36] N. Rohland, M. Hofreiter, Ancient dna extraction from bones and teeth, Nat. Protoc. 2 (2007) 1756–1762, https://doi.org/10.1038/nprot.2007.247.
[37] F. Welker, M.J. Collins, J.A. Thomas, M. Wadsley, S. Brace, E. Cappellini, S. T. Turvey, M. Reguero, J.N. Gelfo, A. Kramarz, J. Burger, J. Thomas-Oates, D.A. Ashford, P.D. Ashton, K. Rowsell, D.M. Porter, B. Kessler, R. Fischer, C. Baessmann, S. Kaspar, J.V. Olsen, P. Kiley, J.A. Elliott, C.D. Kelstrup, V. Mullin, M. Hofreiter, E. Willerslev, J.J. Hublin, L. Orlando, I. Barnes, R.D.E. Macphee, Ancient proteins resolve the evolutionary history of Darwin's South American ungulates, Nature 522 (2015) 81–84, https://doi.org/10.1038/nature14249.
[38] S. Presslee, G.J. Slater, F. Pujos, A.M. Forasiepi, R. Fischer, K. Molloy, M. Mackie, J.V. Olsen, A. Kramarz, M. Taglioretti, F. Scaglia, M. Lezcano, J.L. Lanata, J. Southon, R. Feranec, J. Bloch, A. Hajduk, F.M. Martin, R. Salas Gismondi, M. Reguero, C. de Muizon, A. Greenwood, B.T. Chait, K. Penkman, M. Collins, R.D.E. MacPhee, Palaeoproteomics resolves sloth relationships, Nat. Ecol. Evol. 3 (2019) 1121–1130, https://doi.org/10.1038/s41559-019-0909-z.