Univesity of Crete
BIOLOGY DEPARTMENT

Τμήμα Χημείας
Πανεπιστήμιο Κρήτης
Department of Chemistry
University of Crete

Design and method development of Palaeoproteomics analysis in modern and ancient human bones.

Konstantina Tsiminikaki

ΙΝΣΤΙΤΟΥΤΟ ΜΟΡΙΑΚΗΣ ΒΙΟΛΟΓΙΑΣ ΚΑΙ ΒΙΟΤΕΧΝΟΛΟΓΙΑΣ

12 Νοεμβρίου 2020

Design and method development of Palaeoproteomics analysis in modern and ancient human bones.

Authored by: Konstantina Tsiminikaki (902) *1

Supervisor: Associate Professor Michael Aivaliotis *2,3

Examination Committee: Professor Kokkinidis Michael *4

Professor Pergantis Spyridon *5

*1 interdisciplinary MSc programme "Protein Biotechnology", Department of Biology, University of Crete, Heraklion, Greece.

*2 Associate Professor, Laboratory of Biological Chemistry, Department of Basic Sciences and Preventive Medicine, Department of Medicine, Auth, Thessaloniki, Greece.

*3 Group Leader, Functional Proteomics and Systems Biology group (FunPATh), Center for Interdisciplinary Research and Innovation (CIRI-AUTH), Balkan Center, Thessaloniki, Greece.

*4 Peer Professor, Biochemistry, Molecular Biology, Cellular and Developmental Biology, Department of Biology, University of Crete, Heraklion, Greece.

*5 Professor, Division of Enviromental & Analytical Chemistry, Department of Chemistry, University of Crete, Heraklion, Greece.

Acknowledgements

I would like to thank my supervisor, Professor Michael Aivaliotis for giving me the opportunity to work in his laboratory. I am grateful for his advice and help through all this time, as well as for the opportunities he gave me to present and share my work. My acknowledgements to my committee, Professor Michael Kokkinidis and Professor Spyridon Pergantis for their supervision.

I would, also, like to express my appreciation to Konstantina Psatha for all the scientific support, her advice and how supportive she was in every one of my steps. Many thanks to Pr. Elena Kranioti for providing all the modern samples for my thesis with great generosity, as well as for being an ongoing advisor and ready to help every time. In addition, I thank Dr. Georgia Orfanoudaki and our lab technician Nikos Kountourakis for their help and support, and the other members of the lab for their amazing collaboration.

Last but not least, I'd like to thank my family, my friends and especially my partner, for their patience, tolerance and emotional support.

Abstract

Proteomics analysis based on mass spectrometry (MS) is a widely known scientific field, involving the large-scale investigation of the structure and function of a wide range of proteins and their post-translational modifications. Mass spectrometry-based proteomics enables the study of fields of palaeoanthropology, bio-archaeology, archaeology and forensic sciences generally. It offers valuable information about ancient proteins, protein pathways and their post-translational modifications, unlike common anthropology techniques or ancient DNA analysis. Palaeoproteomics analysis has offered a variety of protocols on different kind of samples, like art items, teeth, and bones of various organisms, in order to answer evolutionary or dietary questions, for sex estimation, as well as for dating samples. The aim of this current study was to design and develop an MS-based palaeoproteomics methodology for the study of modern and ancient human samples, particularly bones.

A total of 9 fractionated samples was used, 8 relative modern and one ancient sample. Sample selection and preparation, digestion, desalting and nano liquid chromatography-tandem mass spectrometry (nLC-MS/MS) analysis were contacted following an already established protocol with noteworthy modifications.

This study was successful in extracting, digesting, concentrating, and identifying proteins from modern and ancient human samples, as well as identifying unique proteins. It is necessary to proceed to a further bioinformatics analysis, including other proteomes from bacteria or fungi, in order to identify possible contaminants. In addition, there is the need to extend our sample number and it is important to design a standard sampling and cleaning methodology.

Key words: human bones, palaeoproteomics, nLC-MS/MS, ancient proteins

Περίληψη

Η πρωτεϊνωματική ανάλυση βασισμένη στη φασματομετρία μάζας (MS) είναι ένας ευρέως γνωστός επιστημονικός τομέας, που περιλαμβάνει τη μεγάλης κλίμακας διερεύνηση της δομής και της λειτουργίας ενός ευρέος φάσματος πρωτεϊνών και των μετα-μεταφραστικών τροποποιήσεών τους. Η πρωτεϊνωματική ανάλυση που βασίζεται σε φασματομετρία μάζας επιτρέπει τη μελέτη πεδίων παλαιοανθρωπολογίας, βιο-αρχαιολογίας, αρχαιολογίας και εγκληματολογικών επιστημών γενικότερα. Προσφέρει πολύτιμες πληροφορίες για τις αρχαίες πρωτεΐνες, τις πρωτεϊνικές οδούς και τις μετα-μεταφραστικές τροποποιήσεις τους, σε αντίθεση με τις κοινές τεχνικές ανθρωπολογίας ή την ανάλυση αρχαίου DNA. Η παλαιοπρωτεϊνωματική ανάλυση έχει προσφέρει μια ποικιλία πρωτοκόλλων σε διαφορετικά είδη δειγμάτων, όπως αντικείμενα τέχνης, δόντια και οστά διαφόρων οργανισμών, προκειμένου να απαντηθούν εξελικτικές ή διατροφικές ερωτήσεις, για εκτίμηση φύλου καθώς και για χρονολόγηση. Ο στόχος της παρούσας μελέτης είναι να σχεδιάσει και να αναπτύξει μια μεθοδολογία παλαιοπρωτεϊνωματικής ανάλυσης βασισμένη σε MS για τη μελέτη σύγχρονων και αρχαίων ανθρώπινων δειγμάτων, ειδικά οστών.

Χρησιμοποιήθηκαν συνολικά 9 κλασματοποιημένα δείγματα, 8 σχετικά σύγχρονα και ένα αρχαίο δείγμα. Η επιλογή και προετοιμασία του δείγματος, η πέψη, η αφαλάτωση και η ανάλυση υγρής χρωματογραφίας νανοροής υψηλής απόδοσης συζευγμένη με διαδοχική φασματομετρία μάζας (nLC-MS/MS) πραγματοποιήθηκαν με βάση ένα ήδη καθιερωμένο πρωτόκολλο με σημαντικές τροποποιήσεις.

Η παρούσα μελέτη ήταν επιτυχής στην εξαγωγή, την πέψη, τη συγκέντρωση και την ταυτοποίηση πρωτεϊνών από σύγχρονα και αρχαία ανθρώπινα δείγματα, καθώς και στον εντοπισμό μοναδικών πρωτεϊνών. Είναι απαραίτητο να προχωρήσουμε σε μια περαιτέρω βιοπληροφορική ανάλυση, συμπεριλαμβανομένων άλλων πρωτεϊνωμάτων από βακτήρια ή μύκητες, προκειμένου να εντοπιστεί κάποια πιθανή επιμόλυνση. Επιπλέον, υπάρχει η ανάγκη επέκτασης του αριθμού δειγμάτων και είναι σημαντικό να σχεδιάσουμε μια τυπική μεθοδολογία δειγματοληψίας και καθαρισμού.

Λέξεις κλειδιά: ανθρώπινα οστά, παλαιοπρωτεϊνωματική, nLC-MS/MS, αρχαίες πρωτεΐνες

Table of Contents

List of Figures and Tables

Introduction

Proteomics analysis based on mass spectrometry (MS) is a widely known scientific field and the most dynamic section of analytical technologies, involving the large-scale investigation of the structure and function of a wide range of proteins and their post-translational modifications. It offers valuable information about gene expression, biochemical mechanisms and signaling pathways, which are directly applicable in sciences (Gault and McClenaghan, 2009).

MS-based proteomics enables not only the study of cellular mechanisms, but also allows tackling ambitious goals, such as investigating fields of palaeoanthropology, bio-archaeology, archaeology and forensic sciences generally (Hublin et al., 2020; Welker et al., 2019; Jackson and Barkett., 2016; Zhang et al., 2020; Brockbals et al., 2020; Wood et al., 2006). Referring to these fields, the contribution of proteomic analysis is really significant, as it offers valuable information about ancient proteins, protein pathways and their post-translational modifications, leading to deeper knowledge of the sample's nature and age (Procopio et al., 2018; Wilson et al., 2012). This knowledge cannot be fully investigated through other analyses, like common anthropology techniques or ancient DNA analysis, due to contaminations and lack of a completed and appropriate sample (Hansen et al., 2017).

In the last few years, there has been a significant progress in the commonly known "palaeoproteomics" field, including a variety of protocols (Cleland et al., 2012; Hendy et al., 2018; Procopio and Buckley, 2017) and kind of samples, like art items (Lluveras-Tenorio et al., 2017; Orsini et al., 2018; Buckley et al., 2013; Mackie et al., 2018), teeth (Welker et al., 2020) and bones (Cappellini et al., 2012; Buckley et al., 2019; McGrath et al., 2019). In addition, scientists have been investigating palaeoproteomics, in order to answer evolutionary (Welker et al., 2016; Presslee et al., 2019) or dietary questions (Craig et al., 2000; Hendy et al., 2018), for sex estimation (Glendon et al., 2019) as well as for dating samples using protein modifications (van Doorn et al., 2012).

1.1 Proteomics on teeth

There is a variety of publications regarding proteomic analysis on teeth. Tooth is a complicated sample due to its hardness and demanding procedure to extract proteins (Porto et al., 2011). Also, it is a kind of sample with high possibility of contamination (Mai et al., 2020), due to different eating habits and microbial flora of the mouth (Barbieri et al., 2017; Grassl et al., 2016). In addition, scientists have been examining the dental proteome not only to study its content, variety or for evolutionary purposes (Cappellini et al., 2019; Welker et al., 2020), but also for sex estimation, as it has been shown that some sexually dimorphic proteins can be extracted and analyzed successfully (Stewart et al., 2017). For example, it has been found that sex can be defined by investigating amelogenin proteins, AMELX_HUMAN and AMELY_HUMAN, which play an important role in mineralization of enamel (Glendon et al., 2019). Furthermore, teeth correspond to a part of human tissue that can be maintained over time as it cannot be highly affected by taphonomy and time period.

1.2 Proteomics on bone

1.2.1 Bone structure

Bone structure is extremely complex, and it has attracted the attention of many scientists for different reasons, as it constitutes an important part of endocrine, metabolic and heamatological

processes (Singh et al., 2018). Bone tissue is structured by inorganic elements, like calcium, and organic elements, like collagen and proteoglycans (Clarke., 2008). There are two categories of bone tissue, trabecular bone and cortical bone, which represent 20% and 80% of bone tissue, respectively (Singh et al., 2018; Chang et al., 2017). Structurally, trabecular bone consists of about 75% bone marrow and 25% bone, whereas cortical bone consists of 90% bone and 10% pore space (Chang et al., 2017).

1.2.2 Kinds of bones already analyzed and bone proteomics

There are many publications referring to bone proteomics, which have included different parts of bones, like femur (Capellini et al., 2011; Pérez-Martínez et al., 2016), humerus (Bona et al., 2014), tibia (Pérez-Martínez et al., 2016; Schroeter et al., 2016; Procopio et al., 2017) and many others (Wadsworth and Buckley., 2014; Wadsworth et al., 2017; Welker et al., 2017; Sawafuji et al., 2017; Presslee et al., 2019). Moreover, scientists have been investigating proteins from bones for many years, in order to answer questions dealing with proteome (McGrath et al., 2019; Sawafuji et al., 2017) and proteome degradation (Wadsworth and Buckley., 2014; Wadsworth et al., 2017). Also, some publications have included pathology and diseases (Pérez-Martínez et al., 2016; Bona et al., 2014), sampling, method development (Cappellini et al., 2014; Procopio et al., 2017; Hendy et al., 2018) and evolution (Capellini et al., 2011; Welker et al., 2017; Presslee et al., 2019).

There is a wide variety of proteins that have been extracted and identified from bones. The most commonly found proteins include different kinds of collagen, osteomodulin, osteopontin, glycoproteins, ambumin, biglycan, decorin, various growth factors and other (Salmon et al., 2013; Wadsworth et al., 2014; Welker et al., 2016; Le Meillour et al., 2018; Schroeter et al., 2019; Cappellini et al., 2012).

1.3 Species of organisms already analyzed

As previously mentioned, scientists have already analyzed a variety of proteins coming from different species of organisms, either in the same sample or in different samples. Except from human proteins, scientists have found proteins originating from various organisms.

With regards to art proteomics, proteins of *Gallus gallus* and *Bos Taurus* have been mostly identified (Mackie et al., 2018; Buckley et al., 2013; Orsini et al., 2018; Lluveras-Tenorio et al., 2017). In addition, many proteins coming from postherd (Heaton et al., 2009) or other organisms (Cicatiello et al., 2018; Hendy et al., 2018) have been also found.

Referring to teeth proteomics, some publications are dealing with human samples (Barbieri et al., 2017; Welker et al., 2020), whereas many scientists have analyzed bone samples from humans (Salmon et al., 2013; Bona et al., 2014; Welker et al., 2016; Sawafuji et al., 2017), Mammoth (Cappellini et al., 2012), Bovine (Wadsworth and Buckley, 2014), *Gallus gallus* (Schroeter et al., 2016), *Bos Taurus* (Wadsworth et al., 2017), *Sus scrofa* (Procopio et al., 2017), *Moa species* (Schroeter et al., 2019) and others (Welker et al., 2017; Le Meillour et al., 2018; Buckley et al., 2019; Presslee et al., 2019).

1.4 Aim of the study

The aim of this current study was to design and develop an MS-based palaeoproteomics methodology for the study of modern and ancient human samples, mainly bones. Our goal was to set up a completed protocol and methodology in this innovative field with a universal application. Through advanced bioinformatics analysis it will be possible to identify, quantify and visualize mainly bone proteins, protein pathways and their post-translation modifications, facilitating the deeper understanding of the samples' nature and age. Furthermore, our attempt to analyze the ancient proteome profiles could provide indications on ancient human life and lifestyle, which can be very auxiliary not only to archaeological and evolutionary research, but also to investigations dealing with physiology, pathology and diseases.

Materials & Methods

Background

The first task in this research project was to undertake an extensive bibliography review. Most publications included animal bones and art items, whereas research referring to human samples included mostly teeth. Also, there were many studies dealing with experimental issues, sampling information and advice. In addition, the rationale for conducting a detailed bibliography research review was to design a protocol based on what other scientists have done and what is already known and used in our lab, regarding the currently used analytical techniques.

1. *Samples*

For the purposes of this current research, 8 relatively modern and 1 ancient sample were analyzed (Figure 1). Relatively modern samples were about 100 years old, and they were part of ribs from a 200 Human skeletons collection stored at the Forensic Pathology department of the ministry of Justice in Heraklion, Crete. The collection was a kind gift of Assistant Professor Elena Kranioti (Forensic Sciences, Medical department, University of Crete). Ancient samples were granted by the Ancient DNA Laboratory of the Institute of Molecular Biology and Biotechnology. The ancient sample that was analyzed was a *Homo sapiens* skull fragment aged from about 3.200 B.C.E. From this sample, we analyzed in total three separate subsamples, which were handled as different samples; two cleaned and one NOT cleaned.

All the protocol steps and baseline were adopted by Capellini and his colleagues (Capellini et al., 2012) with some modifications. All the solutions and equipment used were dedicated to this project. Furthermore, protective clothing and gloves were carefully changed between each sample to prevent individual contamination. Summary of the protocol followed can be seen in Figure 3.

Figure 1. Pictures 1 to 8 represent modern samples, and Picture 9 shows the ancient sample.

1a. *Sample selection (1 day)*

All parts of ribs and the ancient skull, except the subsample that was analyzed without cleaning, were cleaned using a Dremel with specific material. The Dremel had a tip made from sandpaper, in order to remove peri- and endo-osteal surfaces, due to possible contamination from different organisms' remains or soil. Then, each bone sample was turned into powder or smaller pieces by wrapping it in aluminium foil and smashing it with a hammer. After that, samples were transferred in labeled eppendorfs, weighed and recorded accordingly (Table 1.).

Code	Weight (mg)
01	254.7
02	246.9
03	283.5
04	234.3
05	40.7
06	351.9
07	741.3
08	207
09a.	36.3
09b.	49
09c.	43.1

Table 1. Bone samples and their weight (9a and 9b are cleaned ancient samples and 9c is NOT cleaned).

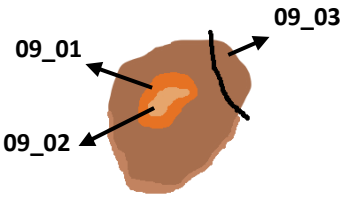


Figure 2. Virtual display of the three different sub-samples taken from the ancient sample (09).

At this point, it is important to mention the difference between the three sub-samples of the ancient sample (09), as shown in Figure 2. 09_01 corresponded to the sub-sample taken after removing the first outer surface of the bone. 09_02 was taken from an inner bone layer, whereas 09_03corresponded to a bone fragment, which did not go through the cleaning process.

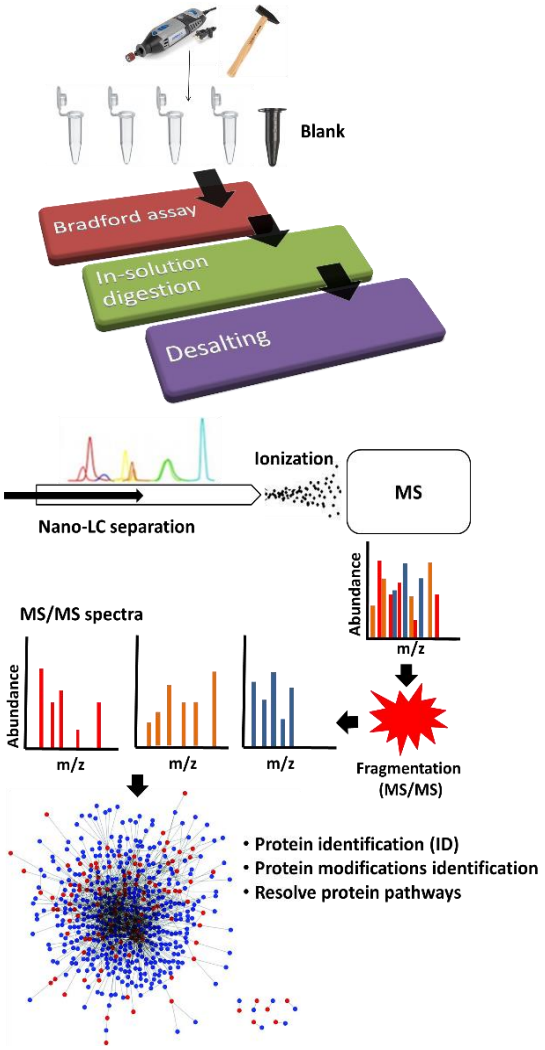


Figure 3. Summary of the methodology followed.

1b. *Sample preparation (4 days)*

After cleaning, each sample was resuspended in 1ml 0.5M EDTA (AppliChem) pH 8.01, in order to remove heavy metals and calcium, and incubated overnight at 4°C with gentle agitation. On the second day, samples were centrifuged for 15 minutes at maximum speed in a bench-top refrigerated centrifuge, at 4°C, discarding the supernatant. Pellets were washed twice with 0.5ml of dH_2O (analytical grade), resuspended in 800μl of 50mM ABS (Merc) pH 7.4 and incubated 24h at 75°C, with gentle agitation. On the third day, samples were centrifuged for 15 minutes at maximum speed, at 4-8°C and supernatant was collected (Fraction A) and stored at -80°C. Pellets were resuspended in 800μl of 50mM ABS pH 7.4 and further incubated for 24h at 75°C, with gentle agitation.

On the fourth and last day of sample preparation, another centrifugation was performed for 15 minutes, at maximum speed, at 4°C. The supernatant (Fraction B) and the pellets (Fraction C) were collected and stored at -80°C, separately.

All different fractions (A, B and C) were collected and stored, in order to be investigated separately.

1c. *Protein concentration (1 day)*

Protein concentration of Fractions A and B was estimated applying the Bradford assay (Bradfrord, 1996; Stoscheck, 1990), using Coomassie (Bradford) Protein Assay Kit (Thermo Fisher) and HITACHI U-1900 UV-Vis Ratio Beam spectrophotometer (Biorad). The assay is based on the observation that the absorbance maximum for an acidic solution of Coomassie Brilliant Blue G-250 shifts from 465 nm to 595 nm when binding to protein occurs. Both hydrophobic and ionic interactions stabilize the anionic form of the dye, causing a visible color change.

At first, standard curve was constructed for protein concentrations determination, ranging from 0.05mg to 1mg (Figure 4), while spectrophotometer was set to 595nm. After measuring the absorbance of all standard dilutions and samples, standard curve and protein concentrations were

estimated using Excel 2019. All modern samples were diluted two times, while ancient samples were not. A volume of solution containing approximately 50μg of proteins was transferred into new appropriately labelled eppendorfs.

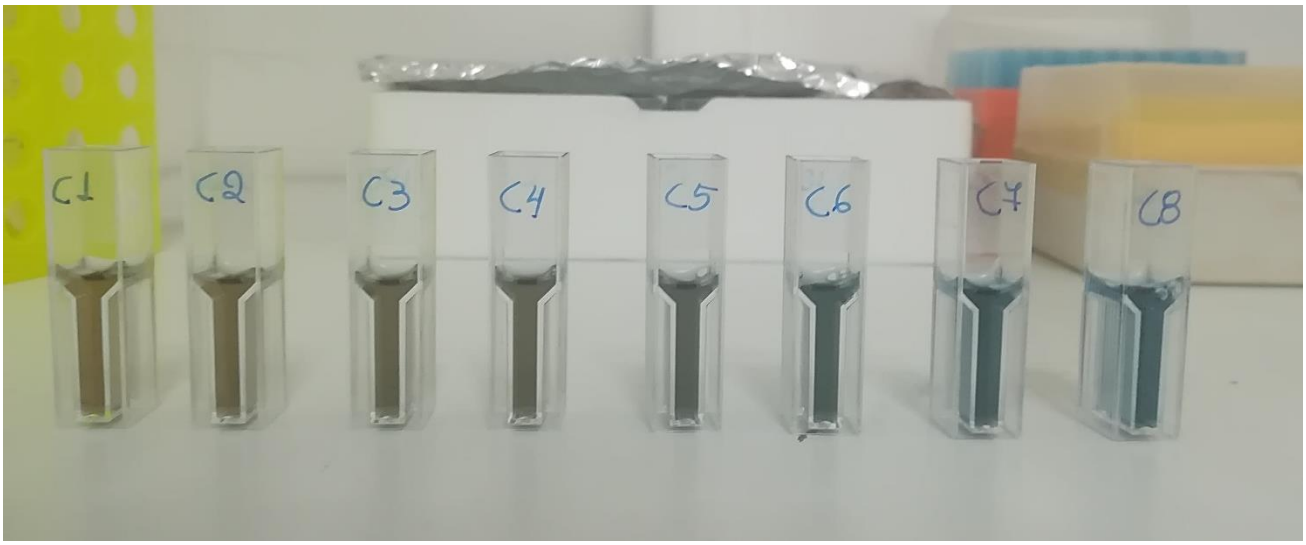


Figure 4. C1 to C8 represent standard dilutions with protein concentrations 0mg, 0.05mg, 0.1mg, 0.2mg, 0.4mg, 0.6mg, 0.8mg and 1mg respectively.

1d. *Preparation and digestion of all fractions (2 days)*

Fraction C (bone pellet) was resuspended in 125μl of 50mM ABS pH 7.4, whereas the volumes used for Fractions A and B were such as to contain 50μg of proteins. Adequate amount of 8M Urea (AppliChem) was added in each tube for protein denaturation, while reducing was succeeded by incubating for 1h at 60°C with 5 mM final concentration of Dithiothreitol (DTT) (AppliChem), prepared immediately before use. After that, alkylation was performed by incubating for 45 min in the dark at room temperature with 15 mM final concentration Iodoacetamide (IAA) (Merc), prepared immediately before use. Finally, the concentration of Urea was reduced with miliQ water below 6M, before adding trypsin for protein digestion, allowing complete trypsinization. After that, 0.5μg/μl of sequencing grade trypsin (ROS) was added and 8% (v/v) final concentration Acetonitrile (ACN) (Sigma Aldrich), while samples were incubated at 37°C with gentle agitation overnight.

The next day, pH was reduced below 2 by adding Trifluoroacetic acid (TFA) (Sigma Aldrich) 0.2-0.8%, checked with pH strips (Merc) and samples were incubated at 37°C for 30 minutes. Then, samples were centrifuged for 15 minutes at maximum speed on a bench-top refrigerated centrifuge at 4°C, and the supernatant was collected in properly labeled Eppendorf tubes discarding the pellet.

1e. *Desalting (1 day)*

In order to remove salt or smaller molecules in each sample, we followed our laboratory's protocol for desalting based on reversed phase chromatography, as shown in Figure 5. The "home-made" C18 columns used for this procedure were loading tips (Kisker Biotech) filled with the help of a pipet tip with a piece of C18 material from a C18 disk (Empore 2215 C18). Column preparation was succeeded using 90% ACN / 5% Formic acid (FA) (AppliChem), while sample preparation was done diluting the sample in 5% FA, sonicating and vortexing it. After sample loading, washing was achieved with 5%

FA, elution of the sample with 90% ACN / 5% FA and lyophilization using a speed-vacuum centrifuge (Thermo Savant) to evaporate the resulted desalted peptide solutions. All samples were stored at 4°C until further processing.

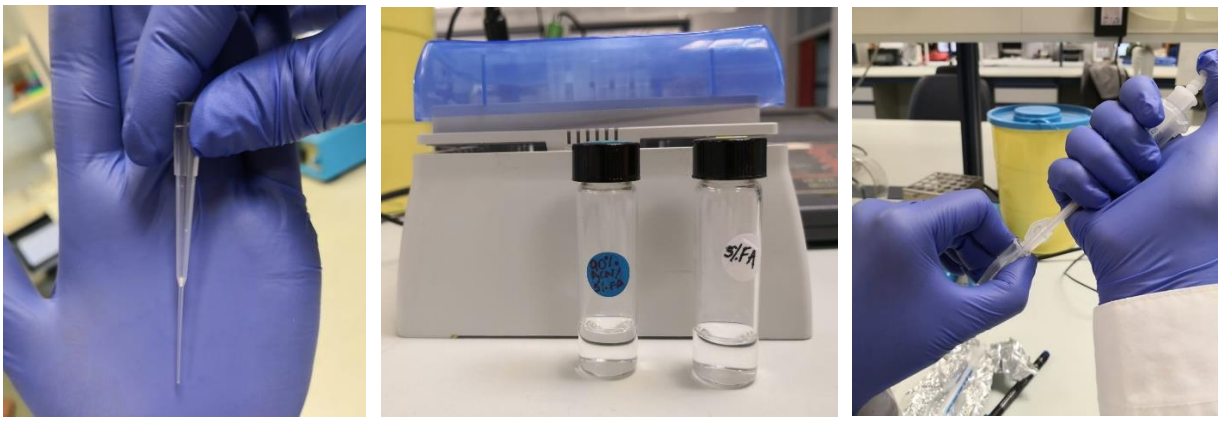


Figure 5. a. Column preparation. b. Mixtures needed. c. Column loading and elution.

1f. *nLC-MS/MS analysis*

Protein identification by nLC-ESI-MS/MS was performed on an LTQ-Orbitrap XL coupled to an Easy nLC (Thermo Scientific, Cheshire, UK). Sample preparation and the LC separation were performed as previously described (Aivaliotis et al., 2007) with minor modifications. For the sample preparation and the nLC-ESI-MS/MS analysis, the dried peptides were dissolved in 5% formic acid aqueous solution, and the tryptic peptide mixtures were separated on a reversed-phase column (Reprosil Pur C18 AQ, Dr. Maisch GmbH), fused silica emitters 100 mm long with a 75µm internal diameter (ThermoFisher Scientific, USA) packed in-house using a packing bomb (Loader kit SP035, Proxeon). The volume of resuspension was equal to the desalting fraction volume with modifications made when needed, especially in samples that were suspicious to obstruct the column (e.g. containing humic substances).

Tryptic peptides were separated and eluted in a linear water-acetonitrile gradient and injected into the MS. Injection volume was set at 3µl, the gradient was set at 3 hours, whereas 2 blank samples of half an hour intervened between each sample. The nLC flow rate was 200 nl min−1. MS survey scans were acquired in the Orbitrap from 200 to 2,000 m/z at a resolution of 60,000, and for the MS/MS, precursor isolation at 1.6 m/z was performed by the quadrupole (Q). Fragmentation of the ten most intense ions by collision induced dissociation (CID) (isolation width, 3 Da; normalized collision energy, 35%; activation q, 0.25; and activation time, 30 ms) in the ion trap analyzer and rapid scan MS analysis were carried out in the ion trap. Each scan included one microscan with a maximum injection time of 200 ms and an automatic gain control sequential mass spectrometry target value of 2×10^4 ions.

1g. *Bioinformatic analysis*

MS data processing

Generated raw data were processed in Proteome Discover (PD v2.2), a software designed for protein identification and quantification in complex biological samples, as well as for post translational modifications' identification (www.thermofisher.com). In order to complete this task, SequestHT search engine (Thermo Scientific) was used based on the complete human proteome from UniProt.

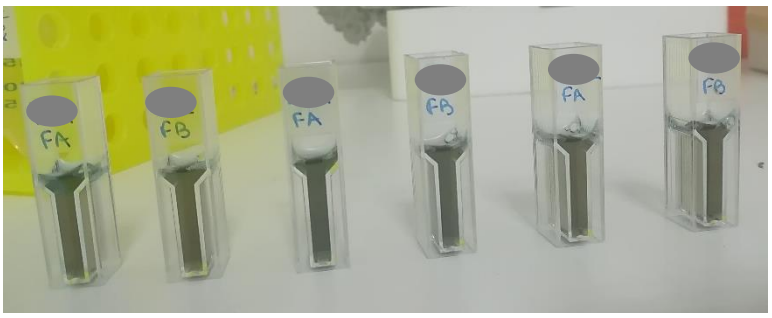
What this algorithm does is matching experimental spectra with theoretical spectra, which have been created from peptide sequences in silico, and then estimating scores to evaluate how well they match (Jiang et al., 2007). Based on these scores, correct and incorrect peptide assignments are determined. The precursor mass tolerance was set at 20 ppm and the fragment ion mass tolerance was 0.8 Da. Spectra were searched for fully tryptic peptides with maximum 3 miss-cleavages. Carbamidomethyl at C was used as static modification. Dynamic modifications included: acetylation of K and peptide N-terminus, oxidation of M and P, deamidation of N and Q. Peptide-spectrum matches (PSMs) confidence was estimated with the Percolator node. PSMs were filtered by setting the maximum Delta Cn parameter of the Percolator node, to 0.01 (maximum Delta Cn is the normalized score difference between the currently selected PSMs and the highest scoring PSM for a spectrum). Peptide FDR was set at 0.01 and validation was based on q-value and decoy database search. Label free quantitation was performed by including the Minora Feature Detection node in the consensus workflow, and the Feature mapper and Precursor Ion quantifier node in the processing workflow. Abundance values were scaled with the "On all average" Scaling Mode.
The PD result files (.msf) were imported in Scaffold 4.9 for further analysis and quantitation. Scaffold 4.9 makes visualization and validation of MS/MS results possible (http://www.proteomesoftware.com/), giving scores for the NSAF, normalized spectral abundance factor, and empai, the exponentially modified protein abundance index, as well as exclusive unique peptide values (McIlwain et al., 2012). NSAF allows the comparison of abundance of proteins in different samples (Zhu et al., 2010), whereas empai is used for absolute quantitation (Yasushi et al., 2005).

Results

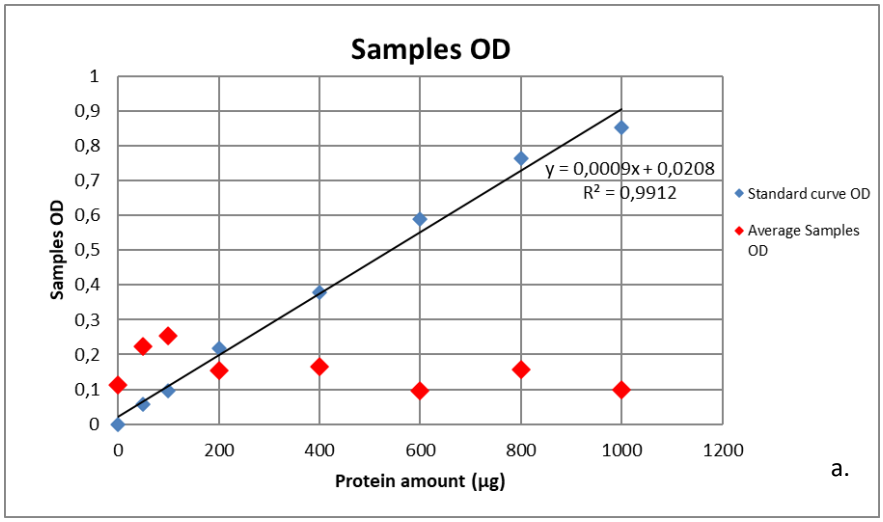
1a. Protein Concentration Determination (Bradford assay)

After constructing standard curve for each batch of Fractions A and B (shown in Figure 3), all samples were measured for protein concentration based on the previously mentioned protocol. A representative picture of the colorimetric measuring of the protein samples is depicted in Picture 1.

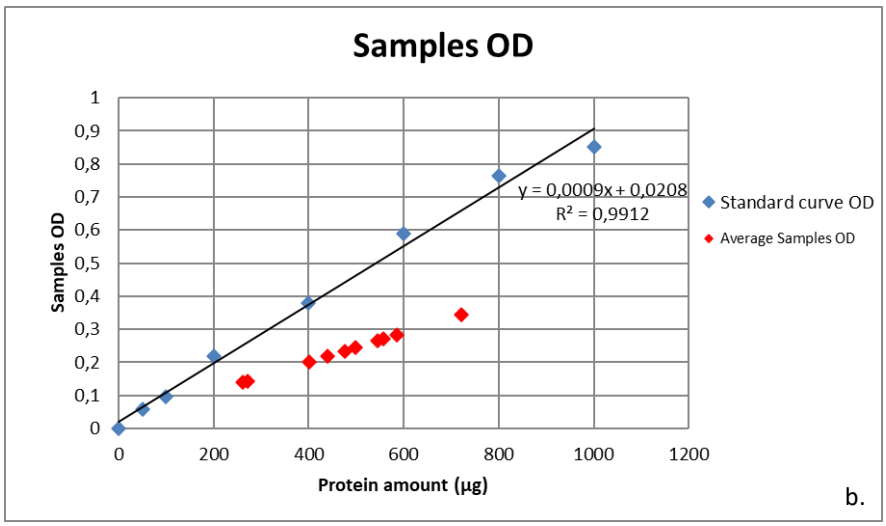
The first ten samples were measured by diluting them 4 times (Figure 6a). All the protein concentrations were at the lowest points of standard curve, therefore the same samples were measured again. This time samples were diluted 2 times and the results are shown in Figure and Table 6b.



Picture 1. A representative picture of a 1 ml Bradford cuvette assay involving Fractions A & B.



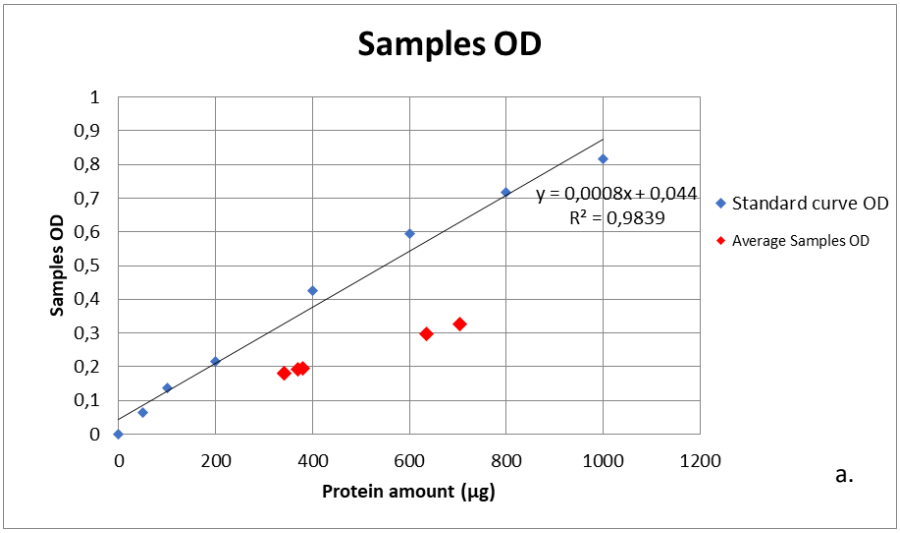
Samples	Concentration (µg/µl)
01_FA	0,41
02_FA	0,90
03_FA	1,03
04_FA	0,59
05_FA	0,64
01_FB	0,33
02_FB	0,60
03_FB	0,34
04_FB	0,35
05_FB	0,22



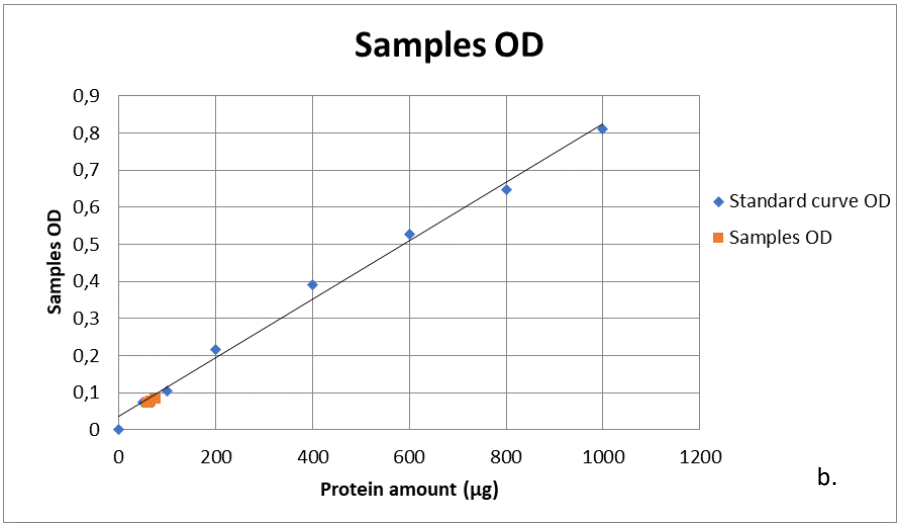
Samples	Concentration (µg/µl)
01_FA	0,44
02_FA	0,72
03_FA	0,54
04_FA	0,56
05_FA	0,48
01_FB	0,27
02_FB	0,50
03_FB	0,59
04_FB	0,40
05_FB	0,26

Figure 6. a. Graphical display of the concentrations measured in the first ten samples (a: DF=4; b: DF=2) and analytical display of the values in the table on the right.

Concentrations for the rest of the modern samples were measured by diluting them two times, and values are shown in Figure 7a. With regards to ancient sample (09), it was handled in three different conditions, as mentioned before; sample taken after cleaning the outer layer (09_01), sample taken from a deeper part of the bone (09_02) and sample NOT cleaned at all (09_03). Protein concetration of 09_01→09_03 was measured by not diluting them, taking into account that protein concentrations in this sample would be low. The results can be seen in Figure 7b.



Samples	Concentration (µg/µl)
06_FA	0,37
06_FB	0,34
07_FA	0,71
07_FB	0,63
08_FA	0,38
08_FB	0,34



Samples	Concentration (µg/µl)
09_01_FA	0,06
09_01_FB	0,08
09_02_FA	0,07
09_02_FB	0,06
09_03_FA	0,08
09_03_FB	0,06

Figure 7. a. Graphical display of the concentrations measured for the rest of the modern samples (DF=2) and analytical display of the values in the table on the right. b. Graphical display of concentrations measured for ancient samples and analytical display of the values in the table on the right.

1b. Desalting

Desalting of all samples was successful after using more than one column for each sample fraction, due to its properties. The resulted lyophilized sample are displayed in Figure 8.

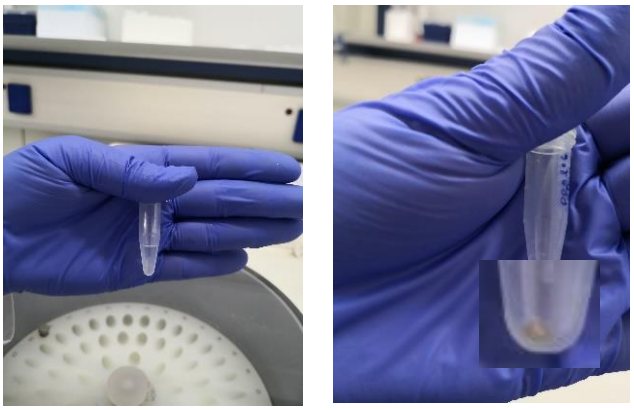
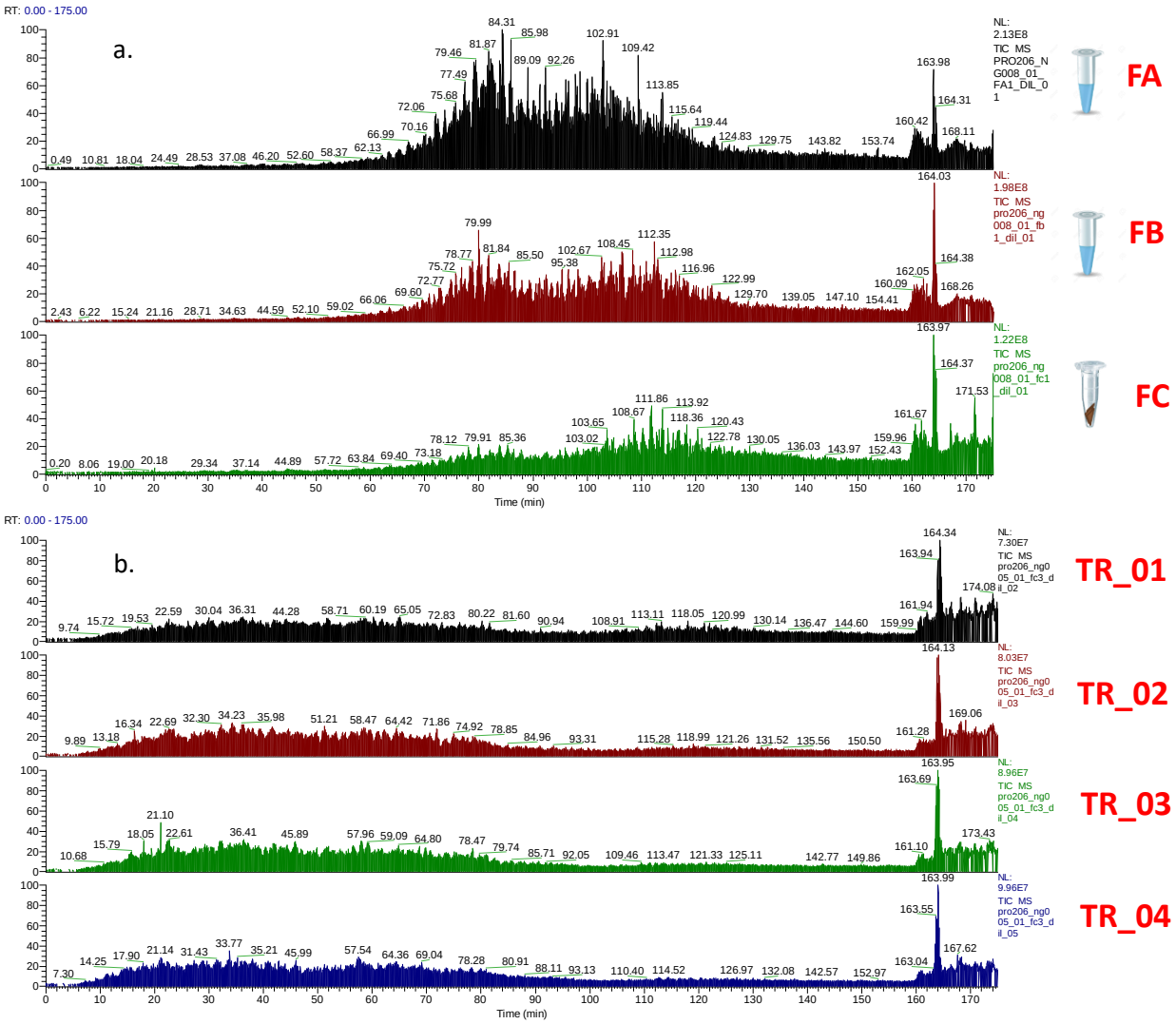


Figure 8. Sample before lyophilization (left picture) and after (right picture).

1c. Mass spectrometry results

Firstly, different dilutions of the samples were tested, as mentioned above, based on the desalting fraction and on the properties of each sample, such as stickiness. A representative chromatograph of all fractions of a sample can be seen in Figure 9a. In addition, the repeatability of the experiment was checked by performing 4 technical repeats for most of the samples. Standard deviation of retention time was about 18%. Finally, the chromatographs of the three different sub-samples, of bone number 09 (ancient sample) based on cleaning steps, are shown in Figure 9c, indicating that they are not the same sample.



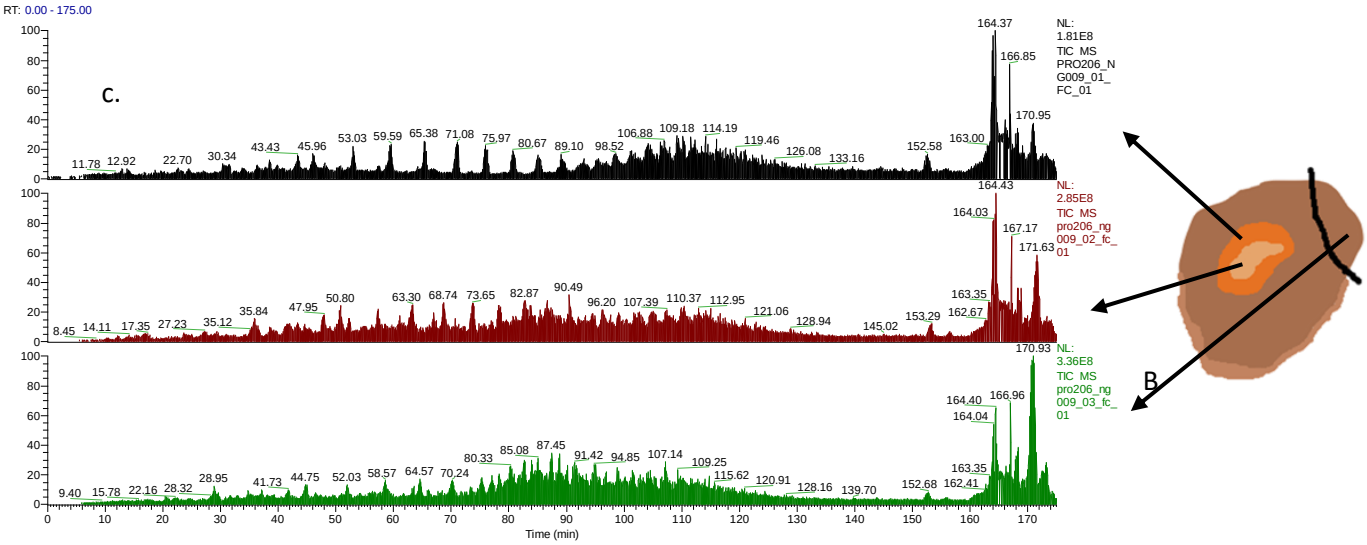


Figure 9. a. Chromatographs of all Fractions of bone 08 in order from FA, FB to FC. b. Chromatographs of technical repeat of bone 05 FC3. c. Chromatographs of Fraction FC of the ancient bone (order: sub-sample of an outer layer, sub-sample of an inner layer and sub-sample of an uncleaned part.

Regarding Bioinformatics analysis, peptide and protein identification was achieved with a possibility of 90-95%, whereas it was possible to identify 2.434 different proteins in modern samples and 1.633 in ancient samples. Among them, the most important ones as they are characteristic for bones, were collagens. In total, 51 proteins, which were either collagens or proteins related to collagens, were identified (Figure 10) having different abundance in each sample. All of them were found mostly in FC fraction, while samples derived from bones 05, 06 and 08 were the most abundant.

For example, collagen alpha-1(I) chain (CO1A1_HUMAN) was found in bone 06 Fraction FC with normalized spectral abundance factor (NSAF) value at 0,00082252 and Exponentially Modified Protein Abundance Index (emPAI) value at 0,083925, which shows the label-free and relative quantitation of the proteins in the sample. Also, protein identification probability for this protein was 100%, while the number of different amino acid sequences that are associated with a single protein group (Exclusive Unique Peptide Count) was 7. On the other hand, collagen alpha-1 (XXIII) chain (CONA1_HUMAN) was identified in bone 05 FC with NSAF value at 0,00070149, emPAI value at 0,16359, 100% protein identification probability and Exclusive Unique Peptide Count at 5.

In addition, collagen alpha-1(XXIV) chain was found in six out of eleven samples, collagen alpha-1(I) chain and collagen alpha-1(XXV) chain were found in five samples, whereas, collagen type XI alpha 2, collagen alpha-1(II) chain, collagen type XVIII alpha 1 isoform CRA d, collagen alpha-6(VI) chain and collagen alpha-6(IV) chain were found in four samples. The rest of the collagen related proteins were discovered in three or two of the samples, while collagen type IV alpha 1 protein variant (Fragment), calcium-binding EGF domain-containing protein 1, collagen type V alpha 3 preproprotein variant (Fragment), collagenase 3 and Alpha 1 type VII collagen variant (Fragment), collagen alpha-1(III) chain, collagen alpha-1(XIII) and many others were found in one sample each.

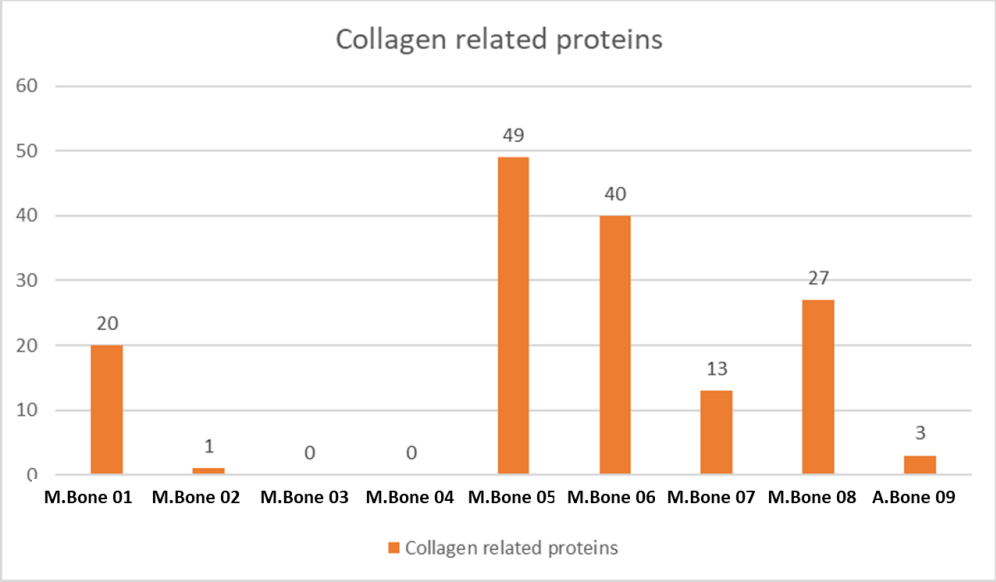


Figure 10. Graphical representation of proteins identified in collagen category.

Regarding to bone-specific proteins, bone marrow stromal antigen 2 was found in samples 05, 06 and 08, bone morphogenetic protein 1 and osteopontin-D (Fragment) in samples 05 and 08 and osteoclast-associated immunoglobulin-like receptor in sample 06. It is important to mention that Cyclin-Y-like protein 1 (A0A494C0K2_HUMAN) was only identified in modern samples 05 and 08, a cyclin related to male sex.

Zooming out from our data, many different kinesins, laminins and keratins were also detected, as well as proteins related to cell energy generation and storage. Most of the proteins were from various categories, such as zinc finger proteins, transcription factors, serum proteins, kinases, myosins, interleucines, while there were several uncharacterized proteins.

In order to compare the protein abundance and relation between the three fractions created for each sample, we constructed venn diagrams for two of the most abundant modern samples, sample number 05 and sample number 08. As it is shown in Figure 11, Fraction C included the most unique proteins, while Fraction A and Fraction B included also proteins not found in the other two Fractions. The same pattern (Figure 12) was noticed in all the Fractions of sample 09 (Ancient bone). Additionally, representative venn diagram of all modern and ancient proteome identified in this research, can be seen in Figure 13.

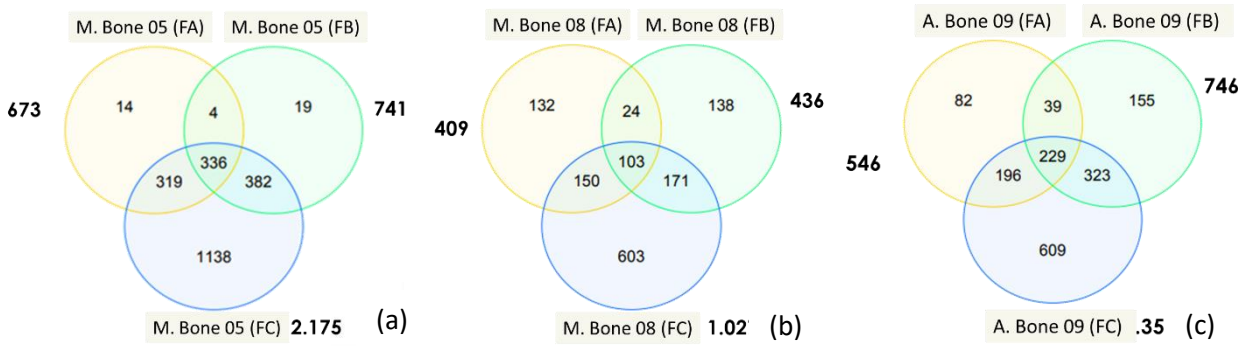


Figure 11. Venn diagrams of all different Fractions created from Modern Bone 05 (a), Modern Bone 08 (b) and Ancient Bone 09 (c).

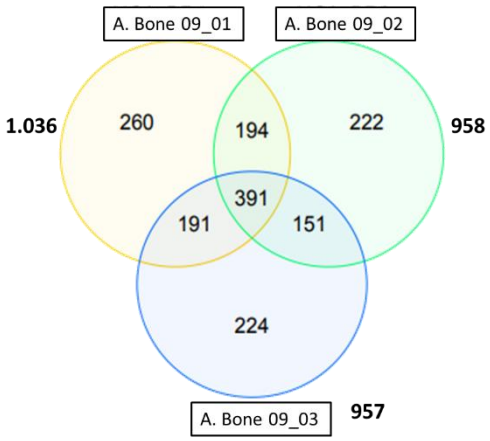


Figure 12. Venn diagram of protein distribution in ancient samples 09_01 09_02 and 09_03.

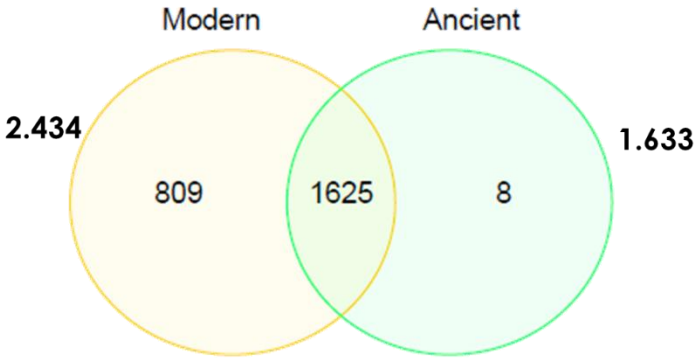


Figure 13. Venn diagram of all modern and ancient proteome found in this research.

More specifically, referring to the ancient bone (sample 09) from which three sub-samples were created, a total of 1633 proteins was identified. The three sub-samples included almost the same number of proteins (Figure 13). There were proteins common to all sub-samples but unique to the ancient sample and some of them are shown in Table 2. Among them, we can find Collagen alpha-6 (IV) chain, Atrial natriuretic peptide-converting enzyme, which is mostly found in muscles (https://www.proteinatlas.org/), and Cilia- and flagella-associated protein 91 which is, also, a male sex-related protein (Uniprot).

Protein Accession [Uniprot]	Protein Name [Uniprot]	Bone 09 01	02	03
Q17RW2	COOA1_HUMAN	98%	100%	98%
A0A1U9X7I9	A0A1U9X7I9_HUMAN	96%	99%	77%
A0A087WZY5	A0A087WZY5_HUMAN	96%	97%	97%
Q5JTC6	AMER1_HUMAN	62%	91%	94%
Q9UGR2	Z3H7B_HUMAN	70%	93%	77%
P0DKV0	S31C1_HUMAN	60%	80%	91%
A0A087X1D5	A0A087X1D5_HUMAN	69%	99%	68%
Q86TB3	ALPK2_HUMAN	81%	99%	55%
Q68DQ2	CRBG3_HUMAN	95%	81%	99%
A8K6J9	A8K6J9_HUMAN	74%	66%	99%
Q9UIG0	BAZ1B_HUMAN	78%	98%	69%
Q7Z4T9	CFA91_HUMAN	88%	98%	74%
B7Z4E9	B7Z4E9_HUMAN	74%	91%	67%
F8VWT9	F8VWT9_HUMAN	82%	70%	97%
A0A0A0MT60	A0A0A0MT60_HUMAN	81%	91%	53%
A8K3Q8	A8K3Q8_HUMAN	81%	77%	91%
B2R9L7	B2R9L7_HUMAN	54%	80%	91%

Table 2. List of proteins found unique and common in all three ancient sub-samples.

Discussion

This study indicates the importance of the establishment of our protocol, as it is the first one to be designed and applied in Greece, while there are only a few laboratories working in this research area all around the world. In addition, this kind of research has a lot to offer in different scientific fields, such as anthropology and evolution (Presslee et al., 2019; Welker et al., 2016), archaeology and zoology (Procopio et al., 2018; Wilson et al., 2012; Martisius et al., 2020; Hublin et al., 2020; Cleland et al., 2016; Buckley et al., 2018), criminology (Procopio et al., 2018) and diseases (Schultz et al., 2007; Nerlich et al., 2018).

Our results showed how successful our method was, developing a protocol for the extraction and identification of proteins and their modifications, originating from an extremely demanding sample. By making improvements into previous protocols and based on our knowledge in this kind of analytical techniques, we were able to extract a sufficient amount of proteins for further analysis. Due to the fact that most of the proteins found in moderate modern/ancient samples are extensively fragmented, it was necessary to extend incubation temperature and time, as previously suggested (Capellini et al., 2012), as well as to concentrate and purify each sample before LC-MS/MS analysis.

Referring to protein concentrations measured by Bradford analysis, it was obvious that concentration was not always analogous to the weight of each sample. More specifically, protein concentration did not increase as the sample became heavier, most of the times. For example, sample 01 was 254.7 mg with total protein concentration of 0.71 μg/μl, whereas sample 02 was 246.9 mg with total protein concentration of 1.22 μg/μl. On the other hand, sample 07, which was the heaviest of all, was 741.3 mg with protein concentration of 1.34 μg/μl, quite similar to sample 02, which was almost three times lighter. All of the above show that the weight of each sample is not indicative of the actual protein concentration, which can be explained through various reasons. First, each bone is like a living organism, thus it can be affected by its environment, temperature, humidity and the organisms that surround it. In addition, handling process during analysis is an important factor, whereas sample's corrosion is a process that depends not only on the age of the individual, but also on the time period that it is buried or environmentally exposed (Emmons et al., 2020).

As regards to the desalting procedure, a significant protocol's step for LC-MS/MS analysis, it was observed that there was a need of using more than one column for each sample. This was unavoidable, especially in cases that protein concentration was high, but, alongside, we noticed that this, also, happened in cases that the sample had a pale-yellow color. Through bibliography research, we found that this color was due to the existence of by-products deriving from microorganisms, especially humic acids (Figure 8, right picture). The lasts are organic material produced by microbial metabolism heterogeneous organic matter produced by microbial metabolism, they are of different molecular sizes and structures, and they are full of aromatic and aliphatic compounds (Qian et al., 2017). In addition, they cause problems in protein purification due to their stickiness, as well as problems in identifying proteins through LC-MS/MS analysis (Arenella et al., 2014). A possible solution to this problem could be the procedure suggested by Schroeter and colleagues, which combines a non-demineralization extraction buffer, concentration of the sample with filters, digestion and a centrifugal stage-tip protocol, reporting promising results (Schroeter et al., 2019).

As seen in Figure 11, fractioning process clearly demonstrated that there was protein loss in each step of the sample preparation, so it is advised to test each Fraction during the analysis, as previously shown (Cappellini et al., 2012). These proteins may be important bone proteins but, on the other hand, it is possible that proteins identified in Fractions A and B may come from other sources, like the soil and plants surrounding the buried sample or microorganisms. In every case, if fractioning is included, it is necessary to confirm the repeatability of the experiment and proceed to protein identification through extensive bioinformatics analysis, to check the possibility of contamination or loss of important sample's proteins.

Bone samples' analysis showed how difficult these samples were in extracting and identifying proteins. Protein abundance was low, and this can be explained by the fact that we only included Human proteome and many modifications. All samples included proteins related to collagens and bone, as well as common human proteins, like keratins and kinesins. It is important to mention that both ancient and modern samples' analysis indicated unique proteins, whereas we were able to identify proteins related to sex. Cyclin-Y-like protein 1 (A0A494C0K2_HUMAN) which was found in modern samples, is related to spermatogenesis (Uniprot). Additionally, Cilia- and flagella-associated protein 91, found in ancient sample, may play a role in spermatogenesis, whereas mutations on this protein's gene are connected to male infertility (https://www.genecards.org/).

At this point, it is important to mention what is shown in Figure 12, regarding all modern and ancient proteome identified. As it was expected, almost the whole ancient proteome was found to be part of the modern proteome, a fact that can be evolutionary explained. More specifically, the age of the ancient sample does not justify a large evolutionary difference within the Human proteome. However, it is significant to proceed to further bioinformatics analysis, in order to learn more information about the unique proteins identified in each proteome.

Besides the above and referring to the ancient bone, it is important to mention the importance of the cleaning process. Sample 09_01 was taken from the bone layer after cleaning the surface, sample 09_02 was from a deeper layer of the same bone and sample 09_03 originated from a completely unprocessed bone-part. The last one, contained many proteins but, in this case, it is almost certain that some of these should not have been identified due to the surrounding environment's effect on the bone (Emmons et al., 2020). Furthermore, ancient bone's protein abundance showed that this sample included more proteins compared to the modern samples, which is confusing considering its inveteracy and the procedures it has suffered, e.g. fossilization. Our hypothesis is that humic acids and fossilization play an important role in this observation, probably due to the fact that humic acids link to proteins, thus preserving them in time, whereas fossilization makes the bone stiffer, thus building a kind of protection of the inside layers. However, no one can dispute the fact that there was protein lost in this sample as well, which can be avoided by following a multi-protease digestion analysis, that uses other enzymes besides trypsin, as it was suggested (Lanigan et al., 2020).

Conclusion

In conclusion, this study was successful in extracting, digesting, concentrating and identifying proteins in modern and ancient human samples, as well as identifying unique proteins characterizing the samples. Taking under consideration possible contamination and the importance of this kind of

research, it is necessary to proceed to a further bioinformatics analysis, including other proteomes from bacteria or fungi, in order to identify possible contaminants. Additionally, there is the need of applying our protocol to more samples, modern and especially ancient, to confirm our results, find common proteins or identify unique proteins, defining a sample. Finally, it is important to design a standard sampling and cleaning methodology, so that extra care will be given to avoid further contamination.

References

Aivaliotis, M. et al. (2007) Large-scale identification of N-terminal peptides in the halophilic archaea Halobacterium salinarum and Natronomonas pharaonis. J Proteome Res; 6: 2195–2204.

Arenella M., Giagnoni L., Masciandaro G., Ceccanti B., Nannipieri P., Renella G. (2014) Interactions between proteins and humic substances affect protein identification by mass spectrometry. *Biology and Fertility of Soils;* 50 (3): 447-454.

Barbieri R, Mekni R, Levasseur A, et al (2017). Paleoproteomics of the Dental Pulp: The plague paradigm. *PLoS One.*; 12(7): e0180552.

Bona A, Papai Z, Maasz G, et al. (2014). Mass spectrometric identification of ancient proteins as potential molecular biomarkers for a 2000-year-old osteogenic sarcoma. *PLoS One.*; 9(7): e103862.

Bradford, MM. (1976) A rapid and sensitive for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Analytical Biochemistry; 72: 248-254.

Brockbals L, Staeheli SN, Kraemer T, Steuer AE. (2020). Postmortem metabolomics: Correlating time-dependent concentration changes of xenobiotic and endogenous compounds [published online ahead of print, 2020 May 5]. *Drug Test Anal*., 10.1002

Buckley M, Lawless C, Rybczynski N. (2019) Collagen sequence analysis of fossil camels, Camelops and c.f. Paracamelus, from the Arctic and sub-Arctic of Plio-Pleistocene North America. *J Proteomics*.; 194: 218-225.

Buckley M, Melton ND, Montgomery J. (2013). Proteomics analysis of ancient food vessel stitching reveals >4000-year-old milk protein. *Rapid Commun Mass Spectrom*.; 27(4): 531-538.

Buckley M, Melton ND, Montgomery J. Proteomics analysis of ancient food vessel stitching reveals >4000-year-old milk protein. (2013). *Rapid Commun Mass Spectrom*. 27(4): 531-538.

Cappellini E, Collins MJ, Gilbert MT. (2014). Biochemistry. Unlocking ancient protein palimpsests. *Science*.; 343(6177): 1320-1322.

Cappellini E, Jensen LJ, Szklarczyk D, et al. (2011). Proteomic analysis of a pleistocene mammoth femur reveals more than one hundred ancient bone proteins. *J Proteome Res*.; 11(2): 917-926.

Cappellini E, Welker F, Pandolfi L, et al. (2019). Early Pleistocene enamel proteome from Dmanisi resolves Stephanorhinus phylogeny. *Nature*; 574(7776): 103-107.

Chang G, Boone S, Martel D, et al. (2017). MRI assessment of bone structure and microarchitecture. *J Magn Reson Imaging*.; 46(2): 323-337.

Cicatiello P, Ntasi G, Rossi M, Marino G, Giardina P, Birolo L. (2018) Minimally Invasive and Portable Method for the Identification of Proteins in Ancient Paintings. *Anal Chem*.; 90(17): 10128-10133.

Clarke B. (2008). Normal bone anatomy and physiology. *Clin J Am Soc Nephrol*; 3(Suppl. 3): S131e9.

Cleland TP, Schroeter ER, Feranec RS, Vashishth D. (2016) Peptide sequences from the first Castoroides ohioensis skull and the utility of old museum collections for palaeoproteomics. *Proc Biol Sci*.; 283(1832): 20160593.

Cleland TP, Voegele K, Schweitzer MH. (2012). Empirical evaluation of bone extraction protocols. *PLoS One*.; 7(2): e31443.

Craig O, Mulville J, Pearson MP, et al. (2000). Detecting milk proteins in ancient pots. *Nature*.; 408 (6810): 312.

Emmons AL, Mundorff AZ, Keenan SW, Davoren J, Andronowski J, Carter DO, DeBruyn JM. (2020) Characterizing the postmortem human bone microbiome from surface-decomposed remains. *PLoS One*; 15(7): e0218636.

Glen P. Jackson, Mark A. Barkett. 2016. History of the Forensic Applications of Mass Spectrometry. Editor(s): Michael L. Gross, Richard M. Caprioli, The Encyclopedia of Mass Spectrometry, *Elsevier*; 271-284.

Glendon J. Parker, Julia M. Yip, Jelmer W. Eerkens, Michelle Salemi, Blythe Durbin-Johnson, Caleb Kiesow, Randall Haas, Jane E. Buikstra, Haagen Klaus, Laura A. Regan, David M. Rocke, Brett S. Phinney. (2019) Sex estimation using sexually dimorphic amelogenin protein fragments in human enamel. *Journal of Archaeological Science.*; 101: 169-180.

Grassl N, Kulak NA, Pichler G, et al. (2016). Ultra-deep and quantitative saliva proteome reveals dynamics of the oral microbiome. *Genome Med.*; 8(1): 44.

Hansen HB, Damgaard PB, Margaryan A, et al. (2017). Comparing Ancient DNA Preservation in Petrous Bone and Tooth Cementum. *PLoS One*.; 12(1): e0170940.

Hendy J, Colonese AC, Franz I, et al. (2018) Ancient proteins from ceramic vessels at Çatalhöyük West reveal the hidden cuisine of early farmers. *Nat Commun.*; 9 (1): 4064.

Hendy J, Welker F, Demarchi B, Speller C, Warinner C, Collins MJ. (2018). A guide to ancient protein studies. *Nat Ecol Evol*.; 2(5): 791-799.

Hublin JJ, Sirakov N, Aldeias V, et al. 2020. Initial Upper Palaeolithic Homo sapiens from Bacho Kiro Cave, Bulgaria. *Nature*; 581 (7808): 299-302.

Jiang, X., Jiang, X., Han, G. et al. (2007) Optimization of filtering criterion for SEQUEST database searching to improve proteome coverage in shotgun proteomics. *BMC* Bioinformatics; 8: 323.

Karl Heaton, Caroline Solazzo, Matthew J. Collins, Jane Thomas-Oates, Edmund T. Bergström. (2009). Towards the application of desorption electrospray ionisation mass spectrometry (DESI-MS) to the analysis of ancient proteins from artefacts. *Journal of Archaeological Science*; 36: 2145-2154

Lanigan LT, Mackie M, Feine S, Hublin JJ, Schmitz RW, Wilcke A, Collins MJ, Cappellini E, Olsen JV, Taurozzi AJ, Welker F. (2020) Multi-protease analysis of Pleistocene bone proteomes. *J Proteomics.*; 228: 103889.

Lluveras-Tenorio A, Vinciguerra R, Galano E, et al. (2017). GC/MS and proteomics to unravel the painting history of the lost Giant Buddhas of Bāmiyān (Afghanistan). *PLoS One*.; 12(4): e0172990.

Louise Le Meillour, Antoine Zazzo, Joséphine Lesur, Sophie Cersoy, Arul Marie, Matthieu Lebon, David Pleurdeau, Séverine Zirah. (2018). Identification of degraded bone and tooth splinters from arid environments using palaeoproteomics. *Palaeogeography, Palaeoclimatology, Palaeoecology.;* 511: 472-482.

Mackie M, Rüther P, Samodova D, et al. (2018). Palaeoproteomic Profiling of Conservation Layers on a 14th Century Italian Wall Painting. *Angew Chem Int Ed Engl*.; 57(25): 7369-7374.

Mai BHA, Drancourt M, Aboudharam G. (2020). Ancient dental pulp: Masterpiece tissue for paleomicrobiology. *Mol Genet Genomic Med*.; 8(6): e1202.

Martisius, N.L., Welker, F., Dogandžić, T. et al. (2020) Non-destructive ZooMS identification reveals strategic bone tool raw material selection by Neandertals. *Sci Rep* 10; 7746.

McGrath K, Rowsell K, Gates St-Pierre C, et al. (2019). Identifying Archaeological Bone via Non-Destructive ZooMS and the Materiality of Symbolic Expression: Examples from Iroquoian Bone Points. *Sci Rep*.; 9(1): 11027.

McIlwain, S., Mathews, M., Bereman, M. S., Rubel, E. W., MacCoss, M. J., & Noble, W. S. (2012). Estimating relative abundances of proteins from shotgun proteomics data. *BMC bioinformatics*, 13, 308.

Montes-Bayón, M. Victor A. Gault, Neville H. McClenaghan. (2009). Understanding bioanalytical chemistry. Principles and applications. *Anal. Bioanal. Chem*.

Nerlich AG. (2018) Molecular paleopathology and paleo-oncology-State of the art, potentials, limitations and perspectives. *Int J Paleopathol*.; 21: 77-82.

Orsini S, Yadav A, Dilillo M, McDonnell LA, Bonaduce I. (2018). Characterization of Degraded Proteins in Paintings Using Bottom-Up Proteomic Approaches: New Strategies for Protein Digestion and Analysis of Data. *Anal Chem*.; 90(11): 6403-6408.

Pérez-Martínez C, Prieto-Bonete G, Pérez-Cárceles MD, Luna A. (2016). Usefulness of protein analysis for detecting pathologies in bone remains. *Forensic Sci Int*. ; 258: 68-73.

Porto IM, Laure HJ, Tykot RH, de Sousa FB, Rosa JC, Gerlach RF. (2011). Recovery and identification of mature enamel proteins in ancient teeth. *Eur J Oral Sci*.; 119 Suppl 1:83-87.

Presslee S, Slater GJ, Pujos F, et al. (2019) Palaeoproteomics resolves sloth relationships. *Nat Ecol Evol*.; 3 (7): 1121-1130.

Procopio N, Buckley M. (2017). Minimizing Laboratory-Induced Decay in Bone Proteomics. *J Proteome Res*.; 16(2): 447-458.

Procopio N, Chamberlain AT, Buckley M. (2017). Intra- and Interskeletal Proteome Variations in Fresh and Buried Bones. *J Proteome Res*.; 16(5): 2016-2029.

Procopio N, Chamberlain AT, Buckley M. (2018). Exploring Biological and Geological Age-related Changes through Variations in Intra- and Intertooth Proteomes of Ancient Dentine. *J Proteome Res*.; 17(3):1000-1013.

Procopio N, Williams A, Chamberlain AT, Buckley M. (2018) Forensic proteomics for the evaluation of the post-mortem decay in bones. *J Proteomics*.; 177: 21-30.

Qian C, Hettich RL. (2017) Optimized Extraction Method To Remove Humic Acid Interferences from Soil Samples Prior to Microbial Proteome Measurements. *J Proteome Res.*; 16(7): 2537-2546.

Salmon CR, Tomazela DM, Ruiz KG, et al. (2013) Proteomic analysis of human dental cementum and alveolar bone. *J Proteomics*.; 91: 544-555.

Sawafuji R, Cappellini E, Nagaoka T, et al. (2017). Proteomic profiling of archaeological human bone. *R Soc Open Sci*.; 4(6): 161004.

Schroeter ER, Blackburn K, Goshe MB, Schweitzer MH. (2019). Proteomic method to extract, concentrate, digest and enrich peptides from fossils with coloured (humic) substances for mass spectrometry analyses. *R Soc Open Sci*.; 6(8): 181433.

Schroeter ER, DeHart CJ, Schweitzer MH, Thomas PM, Kelleher NL. (2016) Bone protein "extractomics": comparing the efficiency of bone protein extractions of *Gallus gallus* in tandem mass spectrometry, with an eye towards paleoproteomics. *PeerJ.*; 4: e2603.

Schultz M, Parzinger H, Posdnjakov DV, Chikisheva TA, Schmidt-Schultz TH. (2007) Oldest known case of metastasizing prostate carcinoma diagnosed in the skeleton of a 2,700-year-old Scythian king from Arzhan (Siberia, Russia). *Int J Cancer*.; 121(12): 2591-5.

Singh S, Bray TJP, Hall-Craggs MA. (2018). Quantifying bone structure, micro-architecture, and pathophysiology with MRI. *Clin Radiol*.; 73(3): 221-230.

Stewart NA, Gerlach RF, Gowland RL, Gron KJ, Montgomery J. (2017). Sex determination of human remains from peptides in tooth enamel. *Proc Natl Acad Sci U S A*; 114(52): 13649-13654.

Stoscheck, CM. (1990) Quantitation of Protein. Methods in Enzymology; 182: 50-69

van Doorn NL, Wilson J, Hollund H, Soressi M, Collins MJ. (2012) Site-specific deamidation of glutamine: a new marker of bone collagen deterioration. *Rapid Commun Mass Spectrom*.; 26 (19): 2319-2327.

Wadsworth C, Buckley M. (2014) Proteome degradation in fossils: investigating the longevity of protein survival in ancient bone. *Rapid Commun Mass Spectrom*.; 28(6): 605-615.

Wadsworth C, Procopio N, Anderung C, et al. (2017). Comparing ancient DNA survival and proteome content in 69 archaeological cattle tooth and bone samples from multiple European sites. *J Proteomics*.; 158: 1-8.

Welker F, Hajdinjak M, Talamo S, et al. (2016) Palaeoproteomic evidence identifies archaic hominins associated with the Châtelperronian at the Grotte du Renne. *Proc Natl Acad Sci U S A*.; 113(40): 11162-11167.

Welker F, Ramos-Madrigal J, Gutenbrunner P, et al. (2020). The dental proteome of Homo antecessor. *Nature*.; 580(7802): 235-238.

Welker F, Smith GM, Hutson JM, et al. (2017) Middle Pleistocene protein sequences from the rhinoceros genus Stephanorhinus and the phylogeny of extant and extinct Middle/Late Pleistocene Rhinocerotidae. *PeerJ*.; 5: e3033.

Welker, F., Ramos-Madrigal, J., Kuhlwilm, M., Liao, W., Gutenbrunner, P., de Manuel, M., Samodova, D., Mackie, M., Allentoft, M. E., Bacon, A. M., Collins, M. J., Cox, J., Lalueza-Fox, C., Olsen, J. V., Demeter, F., Wang, W., Marques-Bonet, T., & Cappellini, E. (2019). Enamel proteome shows that Gigantopithecus was an early diverging pongine. *Nature*; 576(7786), 262–265.

Wilson J, van Doorn NL, Collins MJ. (2012). Assessing the extent of bone degradation using glutamine deamidation in collagen. *Anal Chem*.; 84(21):9041-9048.

Wood M, Laloup M, Samyn N, Mariadel Mar Ramirez Fernandez, Ernst A. de Bruijn, Robert A. A. Maes, Gert De Boeck. (2006). Recent applications of liquid chromatography-mass spectrometry in forensic science*. J Chromatogr A*.; 1130 (1) :3-15.

Yasushi Ishihama, Yoshiya Oda, Tsuyoshi Tabata, Toshitaka Sato, Takeshi Nagasu, Juri Rappsilber, Matthias Mann (2005). Exponentially Modified Protein Abundance Index (emPAI) for Estimation of Absolute Protein Amount in Proteomics by the Number of Sequenced Peptides per Protein. *Molecular & Cellular Proteomics*; 4 (9): 1265-1272.

Zhang Y, Liu L, Ren L. (2020). Liquid chromatography-tandem mass spectrometry (LC-MS/MS) determination of cantharidin in biological specimens and application to postmortem interval estimation in cantharidin poisoning. *Sci Rep*.10(1): 10438.

Zhu, W., Smith, J. W., & Huang, C.-M. (2010). Mass Spectrometry-Based Label-Free Quantitative Proteomics. *Journal of Biomedicine and Biotechnology*; 1–6.