# Human bone paleoproteomics utilizing the SP3 method to maximize detected proteins and reduce humics

Timothy P. Cleland[1]

[1]Museum Conservation Institute, Smithsonian Institution, Suitland, MD 20746

Contact email: clelandtp@si.edu

Abstract

Sample preparation has become an important part of bone proteomics and paleoproteomics and remains one of the major challenges to maximizing the number of proteins characterized from bone extractions. Most paleoproteomic studies have relied on in-solution digestion with the inclusion of filter-aided sample preparation (FASP) as effective methods to detect the proteome. However, neither of these are optimal because few proteins have been detected utilizing only in-solution digestion and the molecular weight cutoff of FASP may miss remaining fragments of proteins in fossil bone. The recently developed single-pot, solid-phase-enhanced sample preparation (SP3) overcomes these issues by not relying on molecular weight while still controlling where the proteins are digested. Here, historical human bones were extracted with either 500 mM tetrasodium EDTA or 400 mM ammonium phosphate dibasic, 200 mM ammonium bicarbonate, 4M guanidine HCl and digested with the SP3 method. Across all samples, 78 ± 7 (400-200-4) and 79 ± 17 (EDTA) protein accessions were identified, including previously difficult to detect proteins such as osteopontin. SP3 also effectively removed 90% or more of the coextracting humic substances (based on reduced absorbance) from extracted proteins. The utility of SP3 for maximizing the number of protein detections in historical bones is promising for future paleoproteomic studies.

Keywords: SP3, bone, human, EDTA, paleoproteomics, sample preparation, osteopontin, collagen I, humics, femur

Introduction

Sample preparation remains one of the most challenging aspects in bone proteomics and paleoproteomics because the hydroxyapatite mineralization necessitates special considerations that other tissue types do not require to access the proteome. Recent exhaustive studies have characterized the extractome of bone protein extraction methodology;[1, 2] however, few post-extraction preparation methods have been used and/or evaluated. Many bone proteomics studies have relied on the use of dialysis approaches[3-9] to remove the demineralization and/or solubilization buffers. Dialysis is relatively slow (i.e., can take days to complete) and is inefficient at removing one of the most commonly utilized demineralizing agents, ethylenediaminetetraacetic acid (EDTA).[10] Alternatively, protein precipitation has had limited usage to remove salts before digestion for bone proteomics.[1, 11] Precipitation has been used less because of protein loss, and difficulties resolubilizing the precipitated proteins,[12] especially for the large amount of collagen I in bone protein extracts. To overcome the issues of dialysis and protein precipitation, buffer exchange and concentration using molecular weight spin filters has become an attractive alternative. An additional benefit to this approach is the ability to digest directly on the filter (i.e., filter-aided sample preparation [FASP][13]) after removal of extraction buffer. FASP has been shown to be effective at detecting bone and other mineralized proteomes.[14] For paleoproteomics, the FASP method has been refined by reducing the molecular weight cutoff from 30,000 to 10,000 to reduce loss of proteins that have been fragmented during diagenesis.[15, 16] Despite its shown effectiveness, FASP is likely not an ideal approach for paleoproteomics because it will lose the smallest protein fragments remaining from the most degraded paleoproteomic samples and has been shown to be less effective for small amounts of protein (e.g., 1 μg).[17] Three additional approaches have been applied in paleoproteomics including: direct protein solubilization and digestion,[18] direct digestion of demineralized bone,[19] and a digestion-free approach.[20] The direct protein solubilization and digestion

approach does not use any additional cleanup prior to digestion reducing the likelihood of protein loss. With direct digestion of demineralized bone, access to insoluble components, including highly glycosylated collagens, becomes possible leading to an increase in protein identifications. Lastly, recent applications of a digestion-free approach has allowed detection of proteins into deep time. This method relies on protein fragmentation as a response to diagenesis and reduces the opportunity to detect competing contamination (e.g., keratin). Despite the directness of these three methods, none reduce or remove contaminating humic substances.

In contrast to the FASP method, a novel paramagnetic bead approach with carboxylate modified surfaces (single-pot solid-phase-enhanced sample preparation [SP3]) has recently been developed[21] and refined[17, 22] for bottom up sample preparation. SP3 has been hypothesized to function in a process similar to hydrophilic interaction liquid chromatography (HILIC) with binding of the proteins in an organic phase and elution in aqueous.[21] On extant proteins, SP3 has been shown to be effective on small amounts of protein[17] as well as in a variety of sample buffers.[21, 22] One major benefit to SP3 over FASP is that it is not molecular weight dependent, so can capture a wide variety of proteins and protein fragments. Additionally, SP3 uses a combination of hydrophobic and hydrophilic beads to maximize the types of proteins that are captured regardless of their hydrophobicity.[17, 21, 22] Here, the SP3 method is applied to historical human bone proteins to evaluate the ability of the method to 1) detect complex proteomes from two different extraction types and 2) remove humic/contaminating brown substances with no additional treatment. Differences in protein composition between the extraction methods are also discussed.

## Materials and Methods

### Human bone and protein extraction

Nine raw, previously powdered human bones (six femora, two tibiae, and one fibula; Table 1) from North America were analyzed. Previously powdered bone was chosen to minimize additional destructive sampling of these human remains. Collection access was granted by Smithsonian National Museum of Natural History Skeletal Biology Program. Between 2.5 and 28 mg of powder were homogenized in 600 µL of either 400 mM ammonium phosphate dibasic, 200 mM ammonium bicarbonate, 4 M guanidine HCl (400-200-4)[4, 23] or 500 mM tetrasodium EDTA using a BeadRuptor Elite (Omni International) for 4 cycles at 6.95 m/s for 45 s with 5 min dwell at -20°C. While not strictly necessary to extend the incubation time of the 400-200-4 method beyond the homogenization cycle,[5] both extraction types were incubated for 66 hr at 37°C with shaking for total incubation consistency. After incubation, both extractions were centrifuged at 18,500 rcf for 10 min then the supernatant decanted and saved. Protein concentration for all supernatants was determined by Pierce Protein Plus (Bradford) Assay. 10 µg of protein from 400-200-4 or 100 µL EDTA were taken for SP3 without any additional treatment.

Table 1: Human bone powders weights used for 400-200-4 and EDTA extraction

| Sample identifier | Burial Date | Bone type | EDTA weight (mg) | 400-200-4 weight (mg) |
|---|---|---|---|---|
| GLO-099-2A | 1862 | Left femur | 15.8 | 15.0 |
| GLO-099-2C | 1862 | Left femur | 21.2 | 18.3 |
| 6CT58-5-AMM03 | 1830-1850 | Left femur | 5.9 | 4.4 |
| 6CT58-5-AMM05 | 1830-1850 | Left femur | 6.3 | 2.5 |
| 6CT58-5-AMM11 | 1830-1850 | Left femur | 16.8 | 17.0 |
| 6CT58-5-AMM13 | 1830-1850 | Right femur | 15.1 | 16.8 |
| 51KEYWORTH-CC-08 | 1850-1899 | Right tibia | 12.7 | 11.3 |
| 7NCE98A-WOODVILLE-06 | 1790-1850 | Right femur | 28.0 | 24.8 |
| 44HE950-RADFORD-5 | 1775-1875 | Left tibia | 7.7 | 6.9 |

SP3

Before binding to the magnetic beads, proteins were reduced and alkylated for 10 min at 95°C in 10 mM TCEP and 40 mM chloroacetamide.[24] Proteins from both extractions were then digested following the refined SP3 method in Moggridge et al.[22] (March 2018 revision from https://github.com/chrishuges/SP3_JPR-2018). Briefly, a 1:1 mixture of hydrophilic and hydrophobic

Sera-Mag Speedbeads (GE Healthcare) were washed three times[21] in 18.3 MΩ MilliQ (Millipore) water with incubation on a 3D printed magnetic stand[25] for two minutes between washes. Beads were stored in MilliQ water until usage. To the reduced and alkylated bone protein solutions, 10 µL of the prepared beads and a volume of ethanol equal to the total volume (sample + bead volume + reducing/alkylating volume) was added and allowed to incubate for 10 min at room temperature. After incubation, beads were placed on the magnetic rack for 2 min to collect the beads and the supernatant was removed and discarded. The beads were subsequently washed three times with 80% ethanol, incubated on the magnetic rack for 2 min, and the supernatant also discarded.

The beads were subsequently resuspended in 90 µL of 50 mM ammonium bicarbonate and 10 µL of 0.04 µg/µL Promega sequencing grade modified trypsin (1:25 trypsin:protein). All samples were incubated overnight at 37°C then peptide supernatants were collected for final C18 stage tip desalting.[26] In short, three 3M Empore C18 disks were cut and placed in 200 µL tips. The tips were wetted with 100% methanol then 80% acetonitrile with 0.1% formic acid, equilibrated with 0.1% formic acid, samples were added, then washed with 0.1% formic acid. All peptides were eluted with 80% acetonitrile with 0.1% formic acid then dried under vacuum. Final peptide samples were resuspended with 0.1% formic acid to a final concentration of 0.5 µg/µL.

Evaluation of humic removal by SP3

A subset of 400-200-4 and EDTA extracts (51KEYWORTH-CC-08, 6CT58-5-AMM13, 7NCE98A-WOODVILLE-06) were treated with SP3 to measure the reduction in potential humic coloration following a protocol similar to Qian and Hettich.[27] Briefly, SP3 was applied and then the pre-treatment, post-SP3 binding, and post-2 hr digestion were measured using a Biotek Epoch microplate spectrophotometer from 250-700 nm. An averaged plate blank was subtracted from all absorbance values and plots were

generated in RStudio 1.1.383 from 300-700 nm to account for interference from the 96-well plate between 250-300 nm.

LC-MS/MS

Peptides (0.5 µg) were trapped on a 3 cm in-house packed Thermo BioBasic C18 column (75 µm i.d., 5 µm particle size), followed by separation on a 20.1 cm in-house-packed Thermo BioBasic C18 column analytical column using a Dionex Ultimate 3000 at 300 nL/min using the following gradient: 0-8 min 2% B, 8-98 min 2-55% B, 98-100 min 55-90% B, 100-103 min 90%B. Buffer A is 0.1% Optima-grade formic acid in Optima-grade water and buffer B is 99.9% Optima-grade acetonitrile with 0.1% Optima-grade formic acid. Peptides were nanosprayed into a ThermoScientific LTQ Orbitrap Velos and analyzed with the following parameters: MS1:60k resolution, 100 ms ion time, 1E6 AGC target; MS2: top 8 peaks fragmented with higher energy collisional dissociation (HCD; 30% NCE), 15k resolution, 5 m/z isolation width, 250 ms AGC, 5E5 AGC target. Randomization for injection order was generated using random.org. EDTA and 400-200-4 samples were randomized together and run in at least triplicate.

Data Analysis

The data analysis was performed using MetaMorpheus version 0.0.285, available at https://github.com/smith-chem-wisc/MetaMorpheus against XML databases of the human proteome (Uniprot; downloaded April 4, 2018) and cRAP database (Uniprot, July 23, 2018). [28-31] The following calibration settings were used: protease = trypsin; maximum missed cleavages = 2; minimum peptide length = 5; maximum peptide length = unspecified; initiator methionine behavior = Variable; fixed modifications = Carbamidomethyl of C, Carbamidomethyl of U; variable modifications = Oxidation of M; max mods per peptide = 3; max modification isoforms = 4096; precursor mass tolerance = 15 ppm; product mass tolerance = 25 ppm. The combined search database contained 20431 non-decoy protein

entries including 230 contaminant sequences. The following G-PTM-D settings were used: protease = trypsin; maximum missed cleavages = 2; minimum peptide length = 5; maximum peptide length = unspecified; initiator methionine behavior = Variable; max modification isoforms = 4096; fixed modifications = Carbamidomethyl of C, Carbamidomethyl of U; variable modifications = Oxidation of M; G-PTM-D modifications count = 335 (including deamidation of NQ and hydroxyproline); precursor mass tolerances = 5 ppm around the masses of the 340 modifications; fragment mass tolerance = 20 ppm. The final database search used the following settings were used: protease = trypsin; maximum missed cleavages = 2; minimum peptide length = 5; maximum peptide length = unspecified; initiator methionine behavior = Variable; fixed modifications = Carbamidomethyl of C, Carbamidomethyl of U; variable modifications = Oxidation of M; max mods per peptide = 3; max modification isoforms = 4096; precursor mass tolerance = 5 ppm; fragment mass tolerance = 20 ppm; report PSM ambiguity = True. All PSMs, peptides, and proteins were filtered at a 1% FDR. Contaminants (keratins and keratin associated proteins, lysozyme c, trypsin), decoy proteins, and peptide standards (cytochrome c digest) were filtered in Microsoft Excel 2016 after the initial FDR filter. Statistical analysis was performed in RStudio 1.1.383 or in Microsoft Excel 2016. For comparison, the dataset from Sawafuji et al. [32] was searched using Metamorpheus version 0.0.286 using similar parameters to the above.

## Results and Discussion

### SP3

Initial SP3 attempts followed the protocols in Hughes et al.[21] and Sielaff et al.[17], but these were found to be suboptimal because of the immiscibility between both 400-200-4 or EDTA buffers and acetonitrile, which required large dilution volumes to produce miscibility (Fig. S1). As a result of the acetonitrile immiscibility, the recently refined ethanol SP3 protocol[22] was applied, and it was found to be effective for both of these buffers without additional dilution (Fig. S1). For the ethanol SP3 protocol, at a 1% FDR

following removal of contaminants and decoys, a total of 101 protein accessions were detected between the two extractions across all samples. A non-significant difference ($p = 0.78$) of 78 ± 7 and 79 ± 17 protein accessions were detected for the nine 400-200-4 and EDTA extracts, respectively. Individually (Table S1, S2, S3; Fig. 1), significantly more protein accessions were detected for the EDTA extraction than for 400-200-4 extraction for GLO-099-2A ($p < 0.01$), GLO-099-2C ($p < 0.05$), and 6CT58-5-AMM11 ($p < 0.05$), and significantly more protein accessions were detected for 400-200-4 than EDTA for 44HE950-RADFORD-5 ($p < 0.001$) and 7NCE98A-WOODVILLE-06 ($p < 0.01$). For target peptides within 1% FDR (Table S1), an average of 471±128 peptides were detected for all nine samples with non-significant differences ($p = 0.29$) of 453 ± 150 and 490 ± 101 peptides detected for 400-200-4 and EDTA, respectively. A significant difference in the number of peptides detected were observed for 6CT58-5-AMM03 ($p < 0.05$), 7NCE98A- WOODVILLE-06 ($p < 0.05$), and 44HE950-RADFORD-5 ($p < 0.0001$) (400-200-4>EDTA) and 6CT58-5-AMM11 ($p < 0.0001$) (EDTA>400-200-4). Before contaminants were removed, 118,329 PSMs, 3975 peptides, and 154 protein groups were detected at a 1% FDR across the 55 raw files.

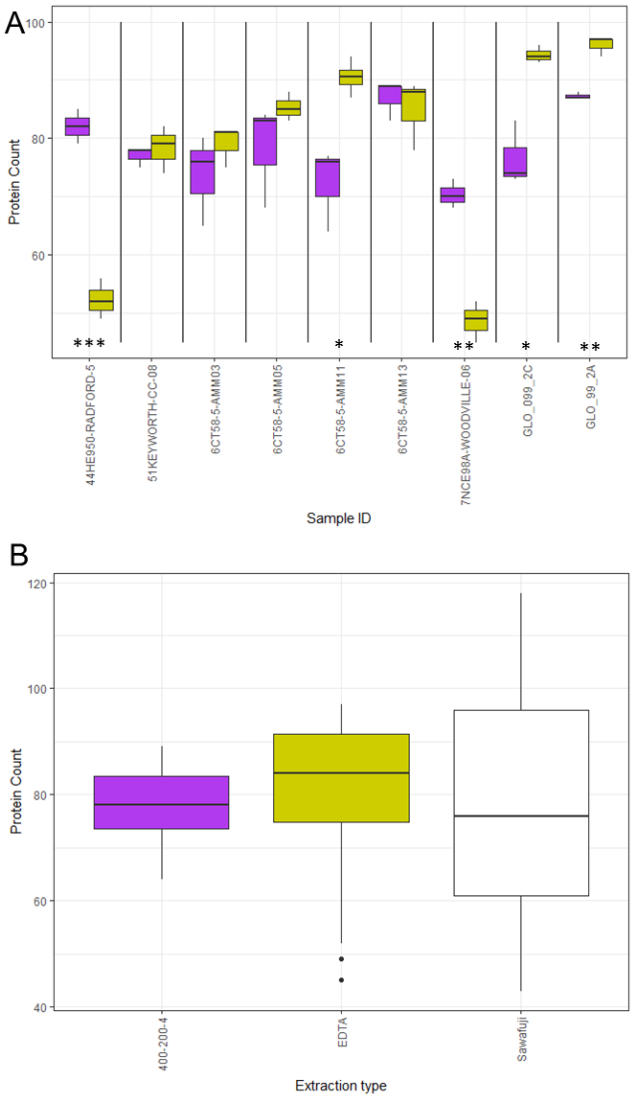



Figure 1: A) Distribution of protein count between samples. Purple is 400-200-4 and yellow is EDTA. *** $p<0.001$; ** $p<0.01$; * $p<0.05$. B) Boxplot comparing protein count between 400-200-4, EDTA, and Sawafuji et al.[32] Boxes indicate the median, first, and third quartile, and the whiskers indicate the minimum and maximum values.

Comparing the results from the human bone extractions here to recently published human paleoproteomes from Japan,[32] insignificant differences for both the peptide count (EDTA: $p = 0.63$; 400-200-4: $p = 0.76$) and protein accessions (EDTA: $p = 0.96$; 400-200-4: $p = 0.97$) were found (Fig. 1B). The closeness in protein count suggests that both EDTA and 400-200-4, combined with SP3, are effective at extracting similar numbers of proteins to the on-pellet digestion used by Sawafuji et al.[32] The diversity

and number of proteins found after SP3 shows that this sample preparation method overcomes previous limitations of both EDTA and 400-200-4. For EDTA, SP3 removes a larger percentage of the salt than dialysis or molecular weight spin columns while reducing the amount of protein loss. The previous applications of 400-200-4 in extant[4] and fossil remains[5] only had limited numbers of protein accessions (extant: 12; fossil: 3) detected after in solution digestion. Again, the major increase in identifications using SP3 on 400-200-4 (6.5-26x than previously reported) may be derived from greater removal of the buffer components or that proteins were lost to the dialysis membrane during processing previously. SP3 provides an environment that is conducive to digestion of these bone protein extracts regardless of the original buffer composition.

Removal of humics

Removal of humic substances that coextract during bone protein extraction has been identified as a challenge in paleoproteomics since at least 2007[7] when they were mentioned as being critical to remove before detection of peptides was possible. As a result of this realization and other mass spectrometry studies showing that humics lead to ion suppression,[27, 33] efficient removal of humics with limited or no protein loss is potentially critical to future paleoproteomic studies, especially as samples become either older or more densely modified by humics. To evaluate reduced humics in these human protein extracts, a subset of the nine samples (i.e., 51KEYWORTH-CC-08, 6CT58-5-AMM13, 7NCE98A-WOODVILLE-06) were chosen to represent the diversity of observed coloration/darkness (Fig. 2A). 51KEYWORTH-CC-08 was visually the lightest (Fig. 2A) of these selected samples, with maximum raw absorbances of 0.27 (EDTA) and 0.15 (400-200-4) at 300 nm. 6CT58-5-AMM13 and 7NCE98A-WOODVILLE-06 (Fig. 2A) were both much darker, and represent the darkest samples in the entire set. 6CT58-5-AMM13 has maximum raw absorbances of 1.2 (EDTA) and 0.97 (400-200-4) and 7NCE98A-WOODVILLE-06 has maximum raw absorbances of 1.4 (EDTA) and 0.85 (400-200-4) at 300 nm. The raw

extracts (Fig. S2, S3) and post-SP3 (Fig. 2B, C) binding supernatants follow similar trends across the wavelengths, while the post-digestion ammonium bicarbonates, regardless of original extraction buffer, are close in intensity to buffer controls (EDTA: 51KEYWORTH-CC-08 4.8%, 6CT58-5-AMM13 2.9%, 7NCE98A-WOODVILLE-06 4.9% of the maximum intensity at 300 nm; 400-200-4: 51KEYWORTH-CC-08 10.3%, 6CT58-5-AMM13 10.5%, 7NCE98A-WOODVILLE-06 9.1%). The 90+% reduction in humics likely explains the higher number of protein identifications than previously shown for EDTA or 400-200-4 because the ion suppression effects of the humics are removed/reduced. Future investigations of the SP3 approach on much darker extractions will show the effectiveness of this method for removing these contaminating molecules and further support that their removal is necessary to characterize the bone paleoproteome. Additionally, unlike the recently proposed precipitation and molecular weight spin filtration to remove humics,[27] SP3 is rapid and separates the protein content from the humics in a single step in a matter of minutes with no required modification of the sample pH beyond the original buffer pH.

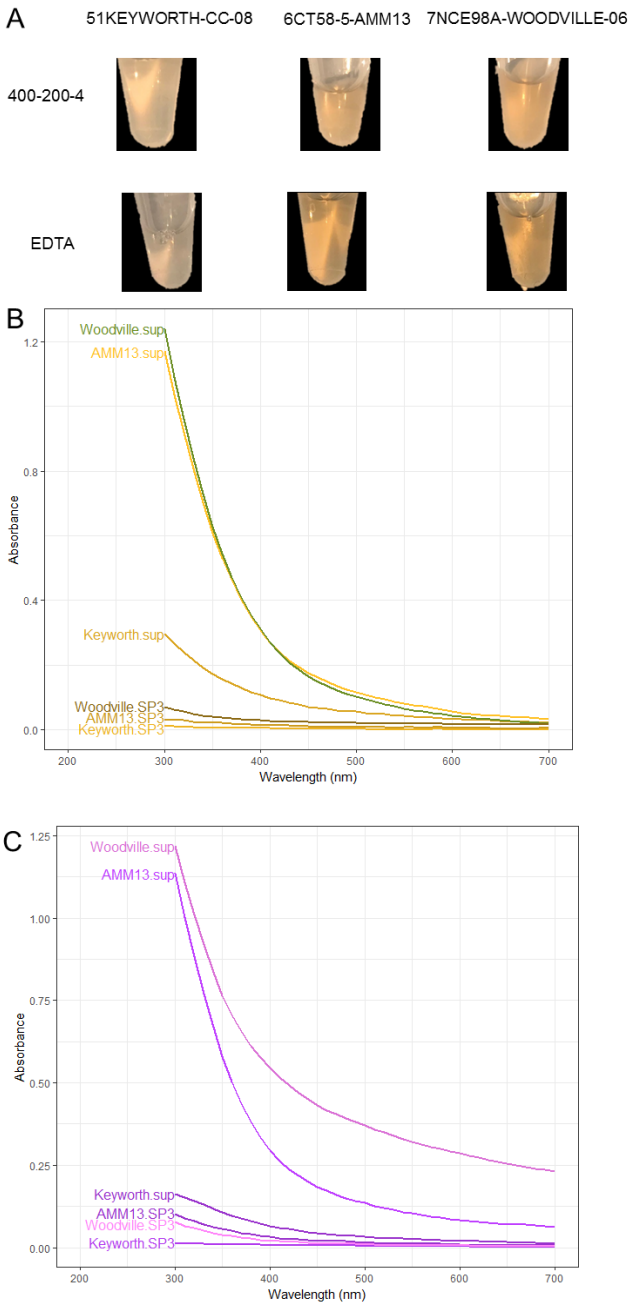



Figure 2: A) Extractions used for evaluation of humic removal showing difference in coloration. Absorbance plots for EDTA (B) and 400-200-4 (C) showing the decrease in humics signal post-digestion. Absorbance curves labeled .sup indicate supernatant after SP3 binding and .SP3 indicates post-digestion absorbance.

Differences in extractomics between buffer types

For clarity with respect to differences between the EDTA and 400-200-4 extraction, results from all individuals were combined regardless of their burial conditions or burial age. Between the two extractions, 43 protein accessions were significantly greater ($p < 0.05$) in intensity for EDTA than 400-200-4 and 20 protein accessions were significantly greater in intensity for 400-200-4 than EDTA ($p < 0.05$). Three main groups of proteins were found to be significantly more concentrated in the EDTA fraction than 400-200-4. These EDTA enriched proteins are: 1) small leucine rich proteoglycans (SLRPs; e.g., lumican, biglycan, decorin) that are associated with collagen fibrillogenesis;[34] 2) mineral-associated and mineralization proteins (e.g., pigment-epithelium derived factor (PEDF), alpha-2-hs-glycoprotein (A2HSG), osteomodulin [OMD], osteopontin [SPP1], SPARC, bone sialoprotein 2 [BSP2][34]); and 3) various growth factors including insulin-like growth factor 1 (IGF-1), its associated binding protein (insulin-like growth factor binding protein 5), insulin-like growth factor 2 (IGF-2), and transforming growth factor beta-1 (TGF-beta 1). In contrast to the EDTA, 400-200-4 has significantly greater relative intensity of collagens I, II, III, and XI. Even though there are significant differences in intensity between the extractions, the majority of the proteins overlap between them (100 of 101 overlap). Consistent with previous bone proteomic and paleoproteomic studies using EDTA and 400-200-4 using other sample preparation methods,[1, 4, 19, 32] SLRPs were detected from all nine samples suggesting that SP3 is also effective for buffer removal and digestion of these collagen associated extracellular proteins.

Despite the increased relative intensity in the EDTA fraction, the mineral-associated proteins were consistently detected in both extraction types reflecting the original hypothesized[4] advantages of the 400-200-4 method and its ability to separate mineral proteins from the hydroxyapatite. Unlike the original description of 400-200-4 using in-solution digestion that was only able to detect osteocalcin from the mineral-associated group, the inclusion of SP3 has allowed for detection of osteopontin from all nine individuals in both the EDTA and 400-200-4 extractions (Fig. 3). Osteopontin is critical for mineralization and microdamage control[35] but has previously been difficult to detect by shotgun

proteomics because it is acidic and highly phosphorylated (as many as 36 phosphorylations[36]). The consistent detection of osteopontin is in contrast to the recent exhaustive extractomics study[1] that was only able to detect it from HCl extraction, but not EDTA. Additional enhancements to the SP3 protocol to include an additional phosphopeptide enrichment will show levels of phosphorylation to osteopontin and other phosphorylated bone proteins, and it will show the levels of lability of phosphorylation in deep time.

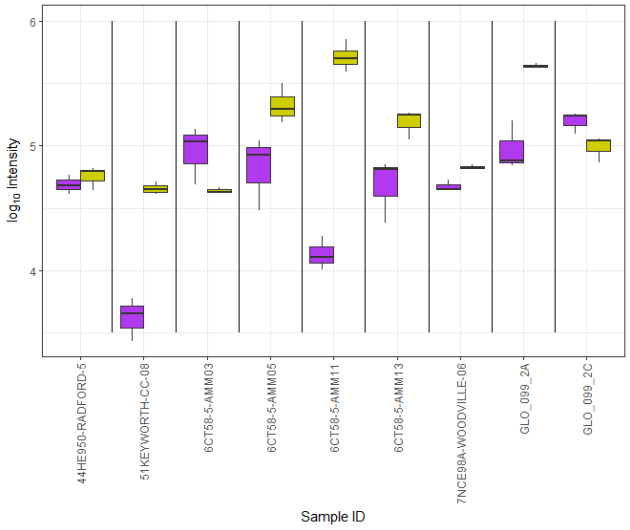



Figure 3: Boxplots of osteopontin intensity by sample and extraction type (Purple boxplots: 400-200-4; Yellow boxplots: EDTA). Boxes indicate the median, first, and third quartile, and the whiskers indicate the minimum and maximum values.

IGF-1, IGF-2, and TGF-beta are associated with bone growth, maintenance, and cell proliferation in bone, and are necessary for long term bone health and remodeling.[37, 38] These growth factors were consistently detected in both the EDTA and 400-200-4 (with limited detection of IGF-1), but previous investigations using in solution for digestion of EDTA[1, 19] or 400-200-4[4] extractions from extant bone were unable to detect these important proteins. This suggests that SP3 allows digestion of proteins that are at a much lower concentration than collagen I that would have less interaction during in solution digestion as a result of collagen I interference (EDTA: Collagen I 75.63% ± 10.08% of TIC, TGF-beta 0.04% ± 0.04%, IGF-2 0.07% ± 0.06%, IGF-1 0.004% ± 0.001%; 400-200-4: Collagen I 82.59% ± 3.8%, TGF-beta

0.01% ± 0.005%, IGF-2 0.01% ± 0.007%, IGF-1 0.0003%, 0.001%). Future assessment of protein concentration using mass spectrometry or other non-mass spectrometry methods between individuals will require the use of only one method, despite the considerable overlap between EDTA and 400-200-4. This may be especially true in assessment of protein-based bone quality that require the examination of particular mineralization proteins.[39-41]

In addition to differences in mineral-associated proteins, SLRPs, growth factors, and collagens with significantly different intensities between EDTA and 400-200-4, periostin and sclerostin were detected in both extractions but were not found to be significantly different in intensity. Periostin and sclerostin are important proteins for mechanical sensing and bone repair and maintenance that have been targets for osteoporotic treatment and fracture risk assessment.[42-45] Despite the similar intensities between EDTA and 400-200-4 extracts, sclerostin was more consistently detected in the EDTA extraction suggesting that a combination of these extraction methods may be required to study these proteins in both extant- and paleoproteomic studies.

To further illustrate the differences between the two extraction methods, a 2-fold difference cutoff was applied, and 37 and 14 protein accessions were found to be both significant ($p < 0.05$) and 2-fold higher in intensity (Fig. 4) for EDTA and 400-200-4, respectively. Like the gross statistical analysis, mineralization proteins, growth factors, and a subset of the SLRPs are more than twice as concentrated in the EDTA extraction. The collagens generally remain after filtering in the 400-200-4, but collagen I alpha 1 does not pass this threshold. This suggests that 400-200-4 preferentially extracts collagen I alpha 2 over collagen I alpha 1. Based on these results, future applications examining the mineralization and mineral-associated proteins would benefit from using EDTA extraction methodology over 400-200-4 to maximize the amount recovered. In contrast, and supporting what was observed by Schroeter et al. that EDTA is less effective than other methods at extracting collagens,[1] 400-200-4 should be used for studies aiming to maximize collagens recovered from bone.

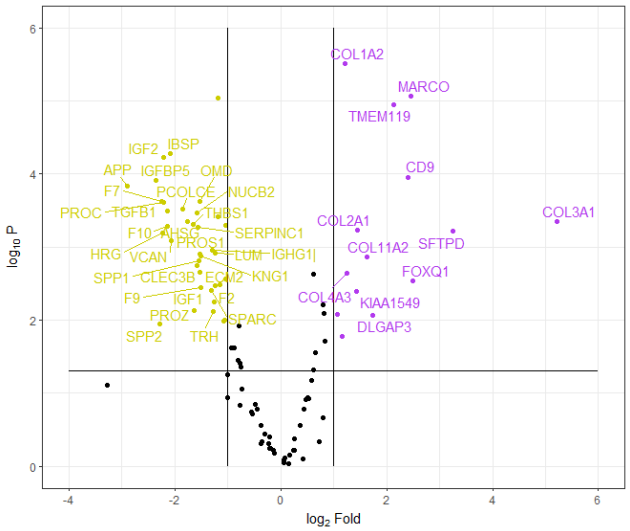



Figure 4: Volcano plot of differences between the EDTA (yellow/left) and 400-200-4 (purple/right) with greater than 2-fold intensity differences.

## Conclusions

SP3 is an effective sample preparation for historical bone that is capable of separating humics while effectively generating peptides from a large diversity of collagenous and non-collagenous bone proteins. Despite this study focusing on historical bone, SP3 is applicable to extant bone proteomics (e.g., protein-based fragility studies, biomarkers) and can result in deep characterization of bone proteomes from a variety of normal and pathological states. Because it efficiently removes EDTA, it is an ideal sample preparation method for bone proteomics and paleoproteomics. While the goal of study was to show the efficacy of SP3 for use in bone proteomics and paleoproteomics, ongoing research will compare SP3 to other sample preparation methods, including FASP. Future applications of this methodology to progressively older bone samples will show how effective this approach is into deep time. Finding a large number of proteins from historical and fossil bone will generate a more complete understanding of the biology and paleobiology of the tested organisms. Possible and those already

realized, but that may benefit from SP3, outcomes of having more proteins include: more complex molecular evolutionary evaluations than morphology or collagen I alone (e.g.,[20]); characterization of diet (e.g., [15, 46]); tracing changes in disease state within ancient populations (e.g., [47]) or across time; or detection of complex post-translational modification changes across species evolution.

## Supporting Information

Table S1: Protein accessions, peptide, and peptide spectral match counts for each sample filtered at 1% FDR; Table S2: Protein groups exported from Metamorpheus; Table S3: Protein groups filtered at 1% FDR with decoys and contaminants removed. Figure S1. Acetonitrile immiscibility between EDTA and 400-200-4. SeraMag beads suspend in the aqueous phase clearly showing the separate layers. Top left to right) EDTA diluted 1x, 2x, 10x, 50x, 100x with water and combined with acetonitrile. Bottom left to right) 400-200-4 diluted 1x, 2x, 10x, 50x, and 100x combined with acetonitrile.

## Acknowledgements

Thanks to Elena Schroeter and Asher Newsome for helpful comments on this manuscript. This project was supported by internal funding at the Smithsonian Institution.

## Data Availability

All raw data, Metamorpheus results, and the R script are available at massIVE (ftp://massive.ucsd.edu/MSV000082826) (Reviewer access: ftp://MSV000082826@massive.ucsd.edu password: boneSP32018

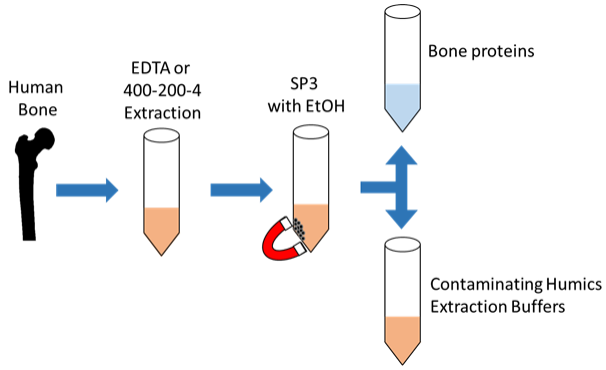
References:

(1) Schroeter, E. R.; DeHart, C. J.; Schweitzer, M. H.; Thomas, P. M.; Kelleher, N. L., Bone protein "extractomics": comparing the efficiency of bone protein extractions of Gallus gallus in tandem mass spectrometry, with an eye towards paleoproteomics. *PeerJ* **2016,** 4, e2603.
(2) Wadsworth, C.; Buckley, M., Characterization of Proteomes Extracted through Collagen-based Stable Isotope and Radiocarbon Dating Methods. *Journal of Proteome Research* **2017,** 17, (1), 429-439.
(3) Cleland, T. P.; Schroeter, E. R.; Schweitzer, M. H., Biologically and diagenetically derived peptide modifications in moa collagens. *Proceedings of the Royal Society B: Biological Sciences* **2015,** 282, (1808), 20150015.
(4) Cleland, T. P.; Vashishth, D., Bone protein extraction without demineralization using principles from hydroxyapatite chromatography. *Analytical Biochemistry* **2015,** 472, (0), 62-66.
(5) Cleland, T. P.; Schroeter, E. R.; Feranec, R. S.; Vashishth, D., Peptide sequences from the first *Castoroides ohioensis* skull and the utility of old museum collections for palaeoproteomics. *Proceedings of the Royal Society of London B: Biological Sciences* **2016,** 283, (1832), 20160593.
(6) Sroga, G. E.; Karim, L.; Colón, W.; Vashishth, D., Biochemical Characterization of Major Bone-Matrix Proteins Using Nanoscale-Size Bone Samples and Proteomics Methodology. *Molecular & Cellular Proteomics* **2011,** 10, (9), M110.006718.
(7) Asara, J. M.; Schweitzer, M. H.; Freimark, L. M.; Phillips, M.; Cantley, L. C., Protein Sequences from Mastodon and Tyrannosaurus Rex Revealed by Mass Spectrometry. *Science* **2007,** 316, (5822), 280-285.
(8) Schweitzer, M. H.; Zheng, W.; Organ, C. L.; Avci, R.; Suo, Z.; Freimark, L. M.; Lebleu, V. S.; Duncan, M. B.; Vander Heiden, M. G.; Neveu, J. M.; Lane, W. S.; Cottrell, J. S.; Horner, J. R.; Cantley, L. C.; Kalluri, R.; Asara, J. M., Biomolecular Characterization and Protein Sequences of the Campanian Hadrosaur *B. canadensis*. *Science* **2009,** 324, (5927), 626-631.
(9) Schroeter, E. R.; DeHart, C. J.; Cleland, T. P.; Zheng, W.; Thomas, P. M.; Kelleher, N. L.; Bern, M.; Schweitzer, M. H., Expansion for the Brachylophosaurus canadensis Collagen I Sequence and Additional Evidence of the Preservation of Cretaceous Protein. *Journal of Proteome Research* **2017,** 16, (2), 920-932.
(10) Mónico, A.; Martínez-Senra, E.; Cañada, F. J.; Zorrilla, S.; Pérez-Sala, D., Drawbacks of Dialysis Procedures for Removal of EDTA. *PLOS ONE* **2017,** 12, (1), e0169843.
(11) Cleland, T. P.; Schroeter, E. R.; Zamdborg, L.; Zheng, W.; Lee, J. E.; Tran, J. C.; Bern, M.; Duncan, M. B.; Lebleu, V. S.; Ahlf, D. R.; Thomas, P. M.; Kalluri, R.; Kelleher, N. L.; Schweitzer, M. H., Mass Spectrometry and Antibody-Based Characterization of Blood Vessels from *Brachylophosaurus canadensis*. *Journal of Proteome Research* **2015,** 14, 5252-5262.
(12) Feist, P.; Hummon, A. B., Proteomic challenges: sample preparation techniques for microgram-quantity protein analysis from biological samples. *Int J Mol Sci* **2015,** 16, (2), 3537-63.
(13) Wisniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M., Universal sample preparation method for proteome analysis. *Nat Methods* **2009,** 6, (5), 359-62.
(14) Warinner, C.; Rodrigues, J. F. M.; Vyas, R.; Trachsel, C.; Shved, N.; Grossmann, J.; Radini, A.; Hancock, Y.; Tito, R. Y.; Fiddyment, S.; Speller, C.; Hendy, J.; Charlton, S.; Luder, H. U.; Salazar-Garcia, D. C.; Eppler, E.; Seiler, R.; Hansen, L. H.; Castruita, J. A. S.; Barkow-Oesterreicher, S.; Teoh, K. Y.; Kelstrup, C. D.; Olsen, J. V.; Nanni, P.; Kawai, T.; Willerslev, E.; von Mering, C.; Lewis, C. M.; Collins, M. J.; Gilbert, M. T. P.; Ruhli, F.; Cappellini, E., Pathogens and host immunity in the ancient human oral cavity. *Nature Genetics* **2014,** 46, (4), 336-+.
(15) Warinner, C.; Hendy, J.; Speller, C.; Cappellini, E.; Fischer, R.; Trachsel, C.; Arneborg, J.; Lynnerup, N.; Craig, O. E.; Swallow, D. M.; Fotakis, A.; Christensen, R. J.; Olsen, J. V.; Liebert, A.; Montalva, N.;

Fiddyment, S.; Charlton, S.; Mackie, M.; Canci, A.; Bouwman, A.; Ruhli, F.; Gilbert, M. T. P.; Collins, M. J., Direct evidence of milk consumption from ancient human dental calculus. *Scientific Reports* **2014,** 4, 1-6.
(16) Cappellini, E.; Gentry, A.; Palkopoulou, E.; Ishida, Y.; Cram, D.; Roos, A. M.; Watson, M.; Johansson, U. S.; Fernholm, B.; Agnelli, P.; Barbagli, F.; Littlewood, D. T. J.; Kelstrup, C. D.; Olsen, J. V.; Lister, A. M.; Roca, A. L.; Dalen, L.; Gilbert, M. T. P., Resolution of the type material of the Asian elephant, *Elephas maximus* Linnaeus, 1758 (Proboscidea, Elephantidae). *Zoological Journal of the Linnean Society* **2014,** 170, (1), 222-232.
(17) Sielaff, M.; Kuharev, J.; Bohn, T.; Hahlbrock, J.; Bopp, T.; Tenzer, S.; Distler, U., Evaluation of FASP, SP3, and iST Protocols for Proteomic Sample Preparation in the Low Microgram Range. *J Proteome Res* **2017,** 16, (11), 4060–4072.
(18) Mackie, M.; Rüther, P.; Samodova, D.; Di Gianvincenzo, F.; Granzotto, C.; Lyon, D.; Peggie, D. A.; Howard, H.; Harrison, L.; Jensen, L. J.; Olsen, J. V.; Cappellini, E., Palaeoproteomic Profiling of Conservation Layers on a 14th Century Italian Wall Painting. *Angewandte Chemie International Edition* **2018,** 57, (25), 7369-7374.
(19) Cleland, T. P., Solid Digestion of Demineralized Bone as a Method To Access Potentially Insoluble Proteins and Post-Translational Modifications. *Journal of Proteome Research* **2018,** 17, (1), 536-542.
(20) Cappellini, E.; Welker, F.; Pandolfi, L.; Ramos Madrigal, J.; Fotakis, A.; Lyon, D.; Moreno Mayar, V. L.; Bukhsianidze, M.; Rakownikow Jersie-Christensen, R.; Mackie, M.; Ginolhac, A.; Ferring, R.; Tappen, M.; Palkopoulou, E.; Samodova, D.; Ruther, P. L.; Dickinson, M. R.; Stafford, T.; Chan, Y. L.; Gotherstrom, A.; Nathan, S. K. S. S.; Heintzman, P. D.; Kapp, J. D.; Kirillova, I.; Moodley, Y.; Agusti, J.; Kahlke, R.-D.; Kiladze, G.; Martinez-Navarro, B.; Liu, S.; Sandoval Velasco, M.; Sinding, M.-H. S.; Kelstrup, C. D.; Allentoft, M. E.; Krogh, A.; Orlando, L.; Penkman, K.; Shapiro, B.; Rook, L.; Dalen, L.; Gilbert, M. T. P.; Olsen, J. V.; Lordkipanidze, D.; Willerslev, E., Early Pleistocene enamel proteome sequences from Dmanisi resolve Stephanorhinus phylogeny. *bioRxiv* **2018**.
(21) Hughes, C. S.; Foehr, S.; Garfield, D. A.; Furlong, E. E.; Steinmetz, L. M.; Krijgsveld, J., Ultrasensitive proteome analysis using paramagnetic bead technology. *Molecular Systems Biology* **2014,** 10, (10).
(22) Moggridge, S.; Sorensen, P. H.; Morin, G. B.; Hughes, C. S., Extending the Compatibility of the SP3 Paramagnetic Bead Processing Approach for Proteomics. *Journal of Proteome Research* **2018**.
(23) Cleland, T. P.; Vashishth, D. Bone Protein Extraction Without Demineralization. US 2017/0369526 A1, Dec. 28, 2017, 2017.
(24) Kulak, N. A.; Pichler, G.; Paron, I.; Nagaraj, N.; Mann, M., Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat Methods* **2014,** 11, (3), 319-24.
(25) acadey Magnetic Rack for 8, 1.5 mL Tubes. https://www.thingiverse.com/thing:79424
(26) Rappsilber, J.; Mann, M.; Ishihama, Y., Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protocols* **2007,** 2, (8), 1896-1906.
(27) Qian, C.; Hettich, R. L., Optimized Extraction Method To Remove Humic Acid Interferences from Soil Samples Prior to Microbial Proteome Measurements. *Journal of Proteome Research* **2017,** 16, (7), 2537-2546.
(28) Solntsev, S. K.; Shortreed, M. R.; Frey, B. L.; Smith, L. M., Enhanced Global Post-translational Modification Discovery with MetaMorpheus. *Journal of Proteome Research* **2018,** 17, (5), 1844-1851.
(29) Li, Q.; Shortreed, M. R.; Wenger, C. D.; Frey, B. L.; Schaffer, L. V.; Scalf, M.; Smith, L. M., Global Post-Translational Modification Discovery. *J Proteome Res* **2017,** 16, (4), 1383-1390.
(30) Shortreed, M. R.; Wenger, C. D.; Frey, B. L.; Sheynkman, G. M.; Scalf, M.; Keller, M. P.; Attie, A. D.; Smith, L. M., Global Identification of Protein Post-translational Modifications in a Single-Pass Database Search. *J Proteome Res* **2015,** 14, (11), 4714-20.

(31) Wenger, C. D.; Coon, J. J., A proteomics search algorithm specifically designed for high-resolution tandem mass spectra. *J Proteome Res* **2013,** 12, (3), 1377-86.
(32) Sawafuji, R.; Cappellini, E.; Nagaoka, T.; Fotakis, A. K.; Jersie-Christensen, R. R.; Olsen, J. V.; Hirata, K.; Ueda, S., Proteomic profiling of archaeological human bone. *Royal Society Open Science* **2017,** 4, (6), 161004.
(33) Piccolo, A.; Spiteller, M., Electrospray ionization mass spectrometry of terrestrial humic substances and their size fractions. *Analytical and Bioanalytical Chemistry* **2003,** 377, (6), 1047-1059.
(34) Boskey, A.; Robey, P. G., The Composition of Bone. In *Primer on the Metabolic Bone Diseases and Disorders of Mineral Metabolism*, Rosen, C. J., Ed. 2013; pp 49-58.
(35) Poundarik, A. A.; Diab, T.; Sroga, G. E.; Ural, A.; Boskey, A. L.; Gundberg, C. M.; Vashishth, D., Dilatational band formation in bone. *Proceedings of the National Academy of Sciences* **2012,** 109, (47), 19178-19183.
(36) Christensen, B.; Nielsen, M. S.; Haselmann, K. F.; Petersen, T. E.; SvТrensen, E. S., Post-translationally modified residues of native human osteopontin are located in clusters: identification of 36 phosphorylation and five O-glycosylation sites and their biological implications. *Biochem. J.* **2005,** 390, (1), 285-292.
(37) Bonewald, L. F.; Mundy, G. R., Role of transforming growth factor-beta in bone remodeling. *Clin Orthop Relat Res* **1990**, (250), 261-76.
(38) Sroga, G. E.; Wu, P. C.; Vashishth, D., Insulin-like growth factor 1, glycation and bone fragility: implications for fracture resistance of bone. *PLoS One* **2015,** 10, (1), e0117046.
(39) Thurner, P. J.; Chen, C. G.; Ionova-Martin, S.; Sun, L.; Harman, A.; Porter, A.; Ager, J. W., 3rd; Ritchie, R. O.; Alliston, T., Osteopontin deficiency increases bone fragility but preserves bone mass. *Bone* **2010,** 46, (6), 1564-73.
(40) Sroga, G. E.; Vashishth, D., Effects of bone matrix proteins on fracture and fragility in osteoporosis. *Curr Osteoporos Rep* **2012,** 10, (2), 141-50.
(41) Tanaka, S.; Narusawa, K.; Onishi, H.; Miura, M.; Hijioka, A.; Kanazawa, Y.; Nishida, S.; Ikeda, S.; Nakamura, T., Lower osteocalcin and osteopontin contents of the femoral head in hip fracture patients than osteoarthritis patients. *Osteoporosis International* **2011,** 22, (2), 587-597.
(42) Bonnet, N.; Garnero, P.; Ferrari, S., Periostin action in bone. *Mol Cell Endocrinol* **2016,** 432, 75-82.
(43) Delgado-Calle, J.; Sato, A. Y.; Bellido, T., Role and mechanism of action of sclerostin in bone. *Bone* **2017,** 96, 29-37.
(44) Ardawi, M. S. M.; Rouzi, A. A.; Al-Sibiani, S. A.; Senani, N. S., Serum Periostin Is Independently Associated with Osteoporosis-Related Fracture Risk among Postmenopausal Women: The Ceor Study. *Osteoporosis International* **2014,** 25, S271-S272.
(45) Rousseau, J. C.; Sornay-Rendu, E.; Bertholon, C.; Chapurlat, R.; Garnero, P., Serum Periostin Is Associated With Fracture Risk in Postmenopausal Women: A 7-Year Prospective Analysis of the OFELY Study. *Journal of Clinical Endocrinology & Metabolism* **2014,** 99, (7), 2533-2539.
(46) Hendy, J.; Warinner, C.; Bouwman, A.; Collins, M. J.; Fiddyment, S.; Fischer, R.; Hagan, R.; Hofman, C. A.; Holst, M.; Chaves, E.; Klaus, L.; Larson, G.; Mackie, M.; McGrath, K.; Mundorff, A. Z.; Radini, A.; Rao, H.; Trachsel, C.; Velsko, I. M.; Speller, C. F., Proteomic evidence of dietary sources in ancient dental calculus. *Proceedings of the Royal Society B-Biological Sciences* **2018,** 285, (1883).
(47) Mackie, M.; Hendy, J.; Lowe, A. D.; Sperduti, A.; Holst, M.; Collins, M. J.; Speller, C. F., Preservation of the metaproteome: variability of protein preservation in ancient dental calculus. *Sci Technol Archaeol Res* **2017,** 3, (1), 74-86.

For TOC Only


Human
Bone
EDTA or
400-200-4
Extraction
SP3
with EtOH
Bone proteins
Contaminating Humics
Extraction Buffers