Article

Proteome Variation with Collagen Yield in Ancient Bone

Noemi Procopio,* Rachel J.A. Hopkins, Virginia L. Harvey, and Michael Buckley*

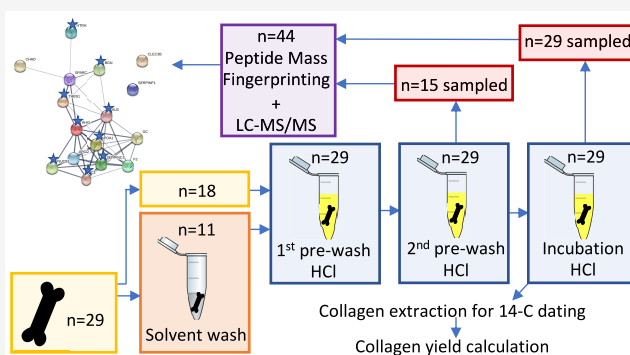Cite This: *J. Proteome Res.* 2021, 20, 1754−1769

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Isotope analyses are some of the most common analytical methods applied to ancient bone, aiding the interpretation of past diets and chronology. For this, the evaluation of "collagen yield" (as defined in radiocarbon dating and stable isotope research) is a routine step that allows for the selection of specimens that are deemed adequate for subsequent analyses, with samples containing less than ~1% "collagen yield" normally being used for isotopic analysis but discounted for radiocarbon dating. The aims of this study were to use proteomic methods of MALDI−TOF (matrix assisted laser desorption ionization time-of-fligh mass spectrometry) and LC−ESI−MS/MS (liquid chromatography electrospray ionization tandem mass spectrometry) to investigate the endogeneity of the dominant proteinaceous biomolecules within samples that are typically considered to contain poorly preserved protein.

Taking 29 archaeological samples, we evaluated the proteome variability between different acid-soluble fractions removed prior to protein gelatinization and considered waste as part of the radiocarbon dating process. We then correlated these proteomes against the commonly used "collagen yield" proxy for preservation. We found that these waste fractions contained a significant amount of both collagenous and noncollagenous proteins (NCPs) but that the abundance of these was not correlated with the acquired "collagen yield". Rather than a depleted protein load as would be expected from a low "collagen yield", the variety of the extracted NCPs was comparable with that commonly obtained from ancient samples and included informative proteins useful for species identification, phylogenetic studies, and potentially even for isotopic analyses, given further method developments. Additionally, we did not observe any correlation between "collagen yield" and peptide mass fingerprint success or between the different fractions taken from the same sample but at different radiocarbon pretreatment stages. Overall, these findings highlight the value in retaining and analyzing sample fractions that are otherwise discarded as waste during the radiocarbon dating process but more importantly, that low "collagen yield" specimens that are often misinterpreted by archaeologists as being devoid of protein can still yield useful molecular sequence-based information.

KEYWORDS: proteomics, ancient bone, collagen, NCPs, radiocarbon dating, stable isotopes

■ INTRODUCTION

The analysis of stable isotopes in bones and teeth is widely used in archaeological and paleontological sciences due to its potential to address questions related to past human activities and ecology.[1−3] In particular, while radiocarbon dating allows the determination of the absolute age of archaeological samples, stable isotope analysis facilitates investigations into past human diet and ecological history.[4] Stable isotopes most widely used for this type of analysis are carbon (δ^{13}C) and nitrogen (δ^{15}N) ratios.[5] This is due to their differential incorporation into the organic and inorganic phases of bones and teeth during their biosynthesis and remodeling[6] that reflects the dietary habits of the specific individual. After death, isotopes are no longer introduced into the tissues, and while radioactive isotopes such as ^{14}C start decaying, the relative concentration of stable isotopes remains constant. Isotopic analyses in bones are most commonly conducted on proteins and in particular, on collagen, the most abundant protein in modern bone, accounting for ~85−90% of the whole bone

proteome.[7] For this reason, the utility of the "collagen yield" (as defined in radiocarbon dating and stable isotope research and sometimes also referred to as "gelatine yield",[8−10] see section "Collagen Yield" in Experimental Section) from archaeological specimens has been the subject of interest for decades. Particularly relevant here is the commonly held belief that samples yielding less than 1% "collagen" should be excluded from radiocarbon dating as a result of their increased susceptibility to contamination from exogenous proteins or carbon sources,[11] which would lead to incorrect data interpretations.
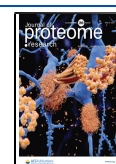
Table 1. Bone Specimens Used in the Study

sample name	cave	approx age[a]	trench	Sq.	layer	depth (cm)	solvent wash	fractions sampled	AG collagen yield (%)	AF collagen yield (%)	morphological identification	ZooMS identification[b,c]	ZooMS + LC−ESI−MS/MS identification[b,c]
Man1	Manastira	Late Holocene	n/a	H24	n/a	n/a		A, B	5.41	4.09	human		human
MR1	Máriaremete	Late Holocene	n/a	6°	sárga agyag	100		A, B	7.25	6.43	human?		human
MR8	Máriaremete	LMP-EUP	n/a	6°	sárga agyag	100		A, B	1.46	1.03			suspected bovidae/cervidae
TD1	Temnata	LMP-EUP	V	O2	5pg	540–545	Y	A, B	0.88	1.08	unidentified mammal	bovine	bovine
TD4	Temnata	LMP-EUP	V	O2	4pg	540–545	Y	A, B	3.10	2.24	unidentified mammal	bovine	bovine
TD5	Temnata	LMP-EUP	V	Λ1	3jx/4	n/a	Y	A, B	2.82	2.22	unidentified mammal		suspected bovidae/cervidae
TD6	Temnata	LMP-EUP	V	H1	3d/4	460–465	Y	A, B	3.62	3.43	unidentified mammal	horse	horse
TD7	Temnata	LMP-EUP	V	Λ2	3jx	505–510	Y	A, B	3.03	3.68	unidentified mammal	horse	horse
TD8	Temnata	LMP-EUP	V	M2	3j/w	480–485	Y	A, B	7.21	5.73	unidentified mammal		suspected bovidae/cervidae
TD14	Temnata	LMP-EUP	V	Λ1	3dg	495–500	Y	A, B	3.94	4.85	unidentified mammal	horse	horse
TD15	Temnata	LMP-EUP	V	Λ2	3dg	480–485	Y	A, B	5.55	6.19	unidentified mammal	horse	horse
TD16	Temnata	LMP-EUP	I	δ2	4	450–452.5	Y	A, B	5.27	3.55	unidentified mammal	horse	horse
TD17	Temnata	LMP-EUP	I	Γ4	4	460–465	Y	A, B	2.57	3.14	unidentified mammal	bovine	bovine
TD19	Temnata	LMP-EUP	V	H1	3d/w	460–465	Y	A, B	3.40	3.50	unidentified mammal	horse	horse
TD20	Temnata	LMP-EUP	V	H1	3d/w	465–470	Y	A, B	7.20	5.28	unidentified mammal	horse	horse
KZ-06	Kozarnika	LMP-EUP	1	E9	4 = IVa	390–393		B	2.73	1.60	large ungulate	cervine	cervine
KZ-10	Kozarnika	LMP-EUP	1	E7	4 = IVa	393–395		B	4.47	3.17	large ungulate	horse	horse
KZ-19	Kozarnika	LMP-EUP	1	E9	4 = IVb	410–415		B	7.57	5.16	bovinae	bovine	bovine
KZ-23	Kozarnika	LMP-EUP	1	D10	4 = IVb	405–410		B	6.17	4.47	large artiodactyla	cervine	cervine
KZ-24	Kozarnika	LMP-EUP	1	F10	5a = V	420–425		B	3.88	2.76	large ungulate	cervine	cervine
KZ-25	Kozarnika	LMP-EUP	1	F10	5a = V	425–430		B	5.85	4.17	large ungulate	horse	horse
KZ-43	Kozarnika	LMP-EUP	1	G5	5c = VII	440–443		B	5.83	3.24	red deer	cervine	cervine
KZ-44	Kozarnika	LMP-EUP	1	F7	6/7 = VIII	460–465		B	0.64		large ungulate	bovine	bovine

Table 1. continued

sample name	cave	approx. age[a]	trench	Sq.	layer	depth (cm)	solvent wash	fractions sampled	AG collagen yield (%)	AF collagen yield (%)	morphological identification	ZooMS identification[b,c]	ZooMS + LC–ESI–MS/MS identification[b,c]
KZ-47	Kozarnika	LMP-EUP	1	F8	5c = VII	445–450		B	4.69	4.59	unidentified mammal		horse
KZ-49	Kozarnika	LMP-EUP	1	G7	5c = VII	450–455		B	3.83	2.68	large ungulate	cervine	cervine
KZ-52	Kozarnika	LMP-EUP	1	F5	6/7 = VIII	453–455		B	4.06	2.76	large ungulate	bovine	bovine
KZ-53	Kozarnika	Late Holocene	3	n/a	n/a	n/a		B	6.07	4.95	human	human	human
KZ-54	Kozarnika	Late Holocene	3	n/a	n/a	n/a		B	2.53		human	human	human
KZ-58	Kozarnika	LMP-EUP	3	F30	6/7? = VIII?	504		B	0.00		unidentified mammal		suspected bovidae/cervidae

Samples that were pretreated with solvents are indicated with "Y" in the "Solvent wash" column. This step was only applied to samples that were suspected of containing exogenous carbon derived from conservation treatment, as is standard procedure for radiocarbon dating.[20] Samples from which the second prewash (HCl) fraction was sampled are indicated by "A", and samples from which the incubation (HCl) fraction was sampled are indicated by "B" in the "Fractions sampled" column (for details see text). "AG" and "AF" collagen yields refer to measurements taken before and after ultrafiltration, respectively. Context information contains, where known, excavation trench ("Trench"), excavation square ("Sq."), and stratigraphic unit ("Layer"). For the latter, at Kozarnika, Arabic numbers indicate geological units, and Roman numerals indicate archaeological units; at Temmata, Arabic numbers were used for geological units, and letters for subunits of varying character, and at Máriaremete, the levels were given descriptions ("sárga agyag" = yellow clay). [a]LMP-EUP = late middle palaeolithic to early upper palaeolithic. [b]Bovine = cattle and bison. [c]Cervine = red deer, fallow deer or elk.

To date, several protocols have been proposed for "collagen" extraction for radiocarbon and stable isotope analysis,[8,12–17] and these pretreatment protocols generally share three main strategies: to demineralize the sample; to remove contaminants such as humic acids, soil contaminants, bone lipids, and exogenous proteins;[18] and finally to solubilize proteins. Several methodological steps are then applied to extract the final "collagen" fraction, and each of these produces liquid fractions that are normally discarded. Previously, Wadsworth and Buckley[19] showed that discarded biomolecular material from these procedures contains numerous noncollagenous proteins (NCPs), with the highest protein variety and abundance being present in the base-soluble fractions obtained from preliminary wash steps performed during the extraction of "collagen" for isotopic or radiocarbon analyses.

The overall process for pretreatment of our samples for collagen extraction at the Oxford Radiocarbon Accelerator Unit (ORAU) consists briefly of the following steps: (1) weighing of the pulverized bone sample, (2) pretreatment of the samples with solvents only if exogenous carbon from conservation/restoration efforts is suspected, (3a) first prewash step with acid, (3b) second prewash step with acid, (3c) incubation step with acid (i.e., demineralization), (3d) cleaning with acid, (4) gelatinization of the sample and extraction of unpurified collagen, (5) freeze-drying of unpurified collagen and weighing (obtaining "after gelatinization" or "AG" collagen yield), (6) collagen hydrolysis and ultrafiltration to obtain purified collagen, and (7) freeze-drying of purified collagen and weighing (obtaining "after filtration" or "AF" collagen yield).[20]

Several studies have been conducted so far in order to find a cheap prescreening methodology for archaeological bones to determine if they could be suitable for radiocarbon dating and isotopic analysis or if they should be discarded from the study.[21] Previous studies proposed the use of attenuated total reflection Fourier transform infrared (ATR–FTIR) spectroscopy as a way to assess the collagen content of bones prior to subjecting them to subsequent disruptive processes such as isotopic,[22] palaeoproteomic,[23–25] and palaeogenetic[26] analyses. Additionally, Harvey et al.[27] proposed the use of Zooarchaeology by Mass Spectrometry (ZooMS) collagen fingerprinting to prescreen bone fragments for radiocarbon dating as a means to evaluate the collagen integrity of bone remains. ZooMS[28] is a peptide fingerprinting method based on mass spectrometric analyses (MALDI–TOF MS) that has been successfully applied in archaeology for species identification of bone fragments too small for conventional (i.e., morphological) identification.[29–32] Harvey et al.[27] found this collagen fingerprinting method to exhibit a 100% success rate with regard to successfully categorizing samples as suitable for dating or not, although further comparisons on a much wider range of specimens from different environments would need to be carried out to validate this. Regardless, this figure is significantly larger than the maximum success rates achieved using %N or C/N investigations (84% and 71%, respectively), which remain the two most commonly used screening techniques for commercial analysis.[33]

Despite the notably greater abundance of collagen in bones in comparison to NCPs, proteomic studies on ancient materials are becoming increasingly popular for clarifying which other proteins survive for prolonged times in archaeological specimens.[34–36] Beyond interpretation of the dominant peptide signals in bone that are largely derived from collagen, a more complex mixture of proteins within a sample can be better analyzed using tandem mass spectrometry (MS/MS), whereby identification is achieved by comparing the collective tandem spectra to sequence databases.[37] NCPs can also be used to obtain phylogenetic information,[38] provide insights into protein diagenetic alternations,[39] as well as on geological[36] and chronological[40] age of specimens and potentially assist in species identification of bone remains.[41] Moreover, NCPs could also provide an approach to performing isotopic analysis on proteins other than collagen. The opportunity to obtain informative NCPs (both for isotopic analyses and for species identification) and diagenetic information on samples normally considered too poorly preserved for radiocarbon dating would be extremely advantageous, especially for precious archaeological artefacts and human remains. Wadsworth and Buckley[34] showed that several NCPs, such as serum albumin, fetuin-A, biglycan, chondroadherin, PEDF, lumican, and prothrombin can be recovered from ancient bone samples up to 900 ka and that in general, the proteome complexity of such samples is inversely proportional to geological age. They also postulated that fetuin-A and albumin might be the most useful NCPs commonly found in ancient samples regardless of their absolute age, proposing that fetuin-A could be the most suitable of all NCPs detected for phylogenetic analysis.[34]

Therefore, the aim of this study was to investigate the survival of NCPs in archaeological specimens, comparing samples classified (by ORAU criteria) as low "collagen yield" samples (hereafter this terminology from radiocarbon dating is preserved for consistency with the literature, despite the fact that the extracted proteome is not purely collagen), with those deemed to have sufficient "collagen yield" for radiocarbon dating. We also aimed to compare the proteomes observed in different fractions obtained from the same sample, focusing on the second prewash and the acid incubation (demineralization) steps with hydrochloric acid, which form part of the routine radiocarbon bone pretreatment protocol at ORAU. Finally, we aimed to investigate limitations of cross-species proteomics in such analyses, investigating differences that may help the taxonomic verification of the samples using LC–ESI–MS/MS for those which generate poor peptide mass fingerprints (via ZooMS),[28] owing to the relatively low abundance of collagen present in the samples.

■ EXPERIMENTAL SECTION

This study was conducted on 29 specimens from four different archaeological sites (Table 1). 14 samples were collected from Kozarnika, a cave in northwestern Bulgaria, 12 were collected from Temnata Dupka, a cave in Bulgaria located about 52 km north of Sofia city, two were collected from Máriaremete (Remete Felső), a cave in Hungary located in the Bakony Mountains, and one sample was collected from Manastira, a cave located in Bulgaria in the Oblast Veliko Tarnovo region.

Protein Extraction

In all 29 cases, the subsamples for proteomic analysis were collected from waste fractions deriving from the radiocarbon dating pretreatment protocol for dating bone collagen.[20]

Bone samples were carefully drilled to collect fine bone powder (410–1170 mg). A sequential solvent wash pretreatment (acetone, methanol, then chloroform [Sigma Aldrich, UK]) was applied to 11 samples, as is customary at ORAU for samples that are suspected to contain exogenous carbon derived from conservation treatment (Table 1 and Figure 1).
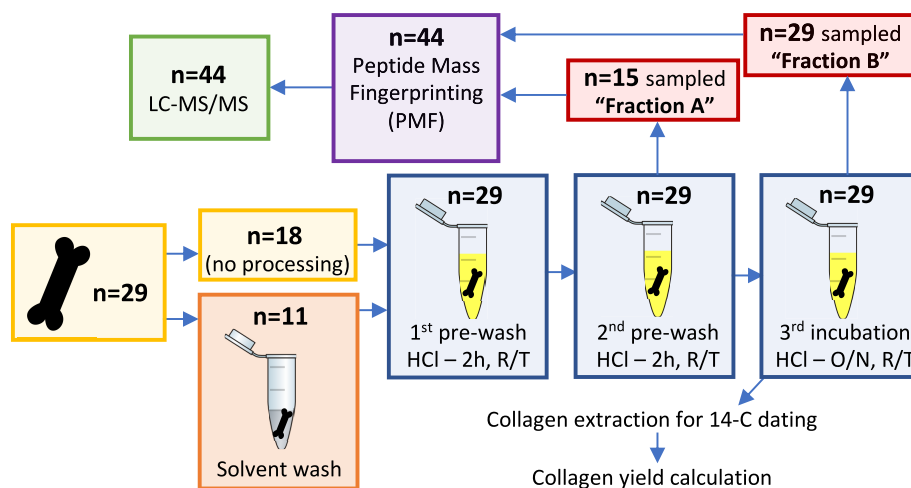
Figure 1. Schematic representation of the treatments and analyses to which samples have been subjected. "R/T" indicates room temperature, "O/N" indicates overnight.

Subsequently, all 29 samples were demineralized with three treatments of 23 mL of 0.5 M hydrochloric acid (HCl) solution (Sigma-Aldrich, UK) and rinsed 3 times with ultrapure water between each replenishment step: (1) 0.5 M HCl (2 h, room temperature [R/T]), (2) 0.5 M HCl (2 h, R/T), and (3) 0.5 M HCl (overnight, R/T). The overnight incubation (*ca.* 18 h) is most comparable to the demineralization step normally applied for ZooMS-type proteomic analyses (e.g. ref 27); therefore, for all samples, 1 mL of solution was collected after the third HCl incubation (labeled fraction B). Additionally, 15 samples had also a 1 mL fraction collected after the second treatment (labeled fraction A) in order to allow for comparison of NCPs between different pretreatment stages (Table 1 and Figure 1), thereby increasing the robustness of our analyses by helping to evaluate whether NPCs are fraction specific. The resulting total of 44 subsamples was split into two fractions each, 0.5 mL for freeze storage (backup) and 0.5 mL for further treatment and analyses (see Table S1 for accession names for LC−ESI−MS/MS analysis and associated sample names in the manuscript).

Samples were ultrafiltrated using 10 kDa molecular weight cut-off filters(MWCO) (Vivaspin, UK) and were buffer exchanged into 50 mM ammonium bicarbonate (ABC, Sigma-Aldrich, UK). Extracted proteins were reduced using 5 mM dithiothreitol (DTT, Sigma-Aldrich, UK) for 40 min at R/T, alkylated with 15 mM iodoacetamide (IAM, Sigma-Aldrich, UK) for 45 min in the dark at R/T, and quenched with a further amount of 5 mM DTT as above. Proteins were then digested with 1 µg of sequencing grade trypsin (Promega, UK) at 37 °C for 5 h. Digestion was stopped by adding 1% trifluoroacetic acid (TFA, Sigma-Aldrich, UK) (to a 0.1% TFA concentration) and then samples were desalted, purified, and concentrated with OMIX C18 reversed-phase Zip-Tips (Agilent Technologies, UK) following manufacturer's protocols. An elution buffer was prepared by mixing acetonitrile (ACN, Sigma-Aldrich, UK) with water and TFA to obtain 50% ACN/0.1% TFA. Peptides were eluted from the Zip-Tips in 100 µL of 50% ACN/0.1% TFA, then samples were dried under a fume cupboard for 1 day, and subsequently resuspended in 20 µL of 5% ACN/0.1% TFA for subsequent MALDI-ToF-MS and LC−ESI−MS/MS analysis.

ZooMS MALDI-ToF-MS Analyses

1 µL of each digest was cocrystalized with 1 µL of 10 mg/mL alpha hydroxycinnamic acid in 50% ACN/0.1% TFA and allowed to air dry on a stainless steel MALDI target plate. Up to 2,000 laser acquisitions were acquired over a *m/z* range 700−3700 following Buckley et al.[28] and compared to the range of megafaunal collagen peptide mass fingerprints acquired for fauna typical of the European Palaeolithic.[42]

LC−ESI−MS/MS Analyses

A total of 44 LC−ESI−MS/MS analyses were performed using an UltiMate 3000 Rapid Separation LC (RSLC, Dionex Corporation, Sunnyvale, CA, USA) coupled to an Orbitrap Elite (Thermo Fisher Scientific, Waltham, MA, USA) mass spectrometer (120 k resolution, full scan, positive mode, normal mass range 350−1500). Peptides were separated on an Ethylene Bridged Hybrid (BEH) C18 analytical column (75 mm × 250 µm i.d., 1.7 µM; Waters) using a gradient from 92% A (0.1% FA in water) and 8% B (0.1% FA in ACN) to 33% B in 44 min at a flow rate of 300 nL min⁻¹. Peptides were then automatically selected for fragmentation by data-dependent analysis; six MS/MS scans (Velos ion trap, product ion scans, rapid scan rate, centroid data; scan event: 500 count minimum signal threshold, top 6) were acquired per cycle, dynamic exclusion was employed, and one repeat scan (i.e., two MS/MS scans total) was acquired in a 30 s repeat duration with that precursor being excluded for the subsequent 30 s (activation: collision-induced dissociation (CID), 2+ default charge state, 2 *m/z* isolation width, 35 eV normalized collision energy, 0.25 activation Q, 10.0 ms activation time).

Proteomic Data Analysis

The collective tandem mass spectra (.mgf) files were then searched against the Swiss-Prot database for matches to primary protein sequences using the Mascot search engine (version 2.5.1; Matrix Science, London, UK), without specific taxonomy filters. Each search included the fixed carbamidomethyl modification of cysteine (+57.02 Da) and the variable modifications for deamidation (asparagine and glutamine, +0.98 Da) and oxidation of lysine, proline, and methionine residues (all +15.99 Da) to account for post-translational modifications and diagenetic alterations (the oxidation of lysine and proline is equivalent to hydroxylation). Enzyme

specificity was selected as trypsin-P (first batch of analyses) and semiTrypsin (second batch of analyses) with one missed cleavage allowed; mass tolerances were set at 5 ppm for the precursor ions and 0.5 Da for the fragment ions. All spectra were considered as having either 2+ or 3+ precursors. Scaffold (v4.10.0, Proteome Software Inc., Portland, OR) was used to validate MS/MS-based peptide and protein identifications. Peptide identifications were accepted if they exceeded specific search engine thresholds (i.e., the suggested peptide homology scores). Protein identifications were accepted if they contained at least 2 identified peptides. Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony. The display option chosen for this work was "total spectrum count" to allow a semiquantitative measurement of the proteins present in the sample. In order to count NCPs, we considered all proteins but excluded from the count collagenous proteins as well as common contaminants and non-intrinsic bone proteins, such as keratins and trypsin. In order to evaluate the presence of peptides that have originated by nonspecific trypsin cleavages, such as those cleaved through the process of diagenesis, an additional Mascot search was performed specifying the digestion enzyme as semiTrypsin. Percentage coverages for specific selected proteins were extracted from Scaffold (display option selected "percent coverage"). RStudio software (version 1.3.959) was used to perform plots and to make statistical analyses using the library tidyverse and the package ggpubr. STRING software version 11.0 was used to calculate functional protein association networks.

"Collagen Yield"

The "collagen yield" for each sample was calculated using data from the radiocarbon dating process, by following standard radiocarbon and stable isotope practices. As a result, we use standard radiocarbon dating terminology in the following description. There were two stages at which "bulk collagen weight" could be measured: before ("AG") and after ultrafiltration ("AF"). More precisely, after the fractions for proteome analyses were collected, the radiocarbon sample was further treated with 0.1 M NaOH (30 min, RT) and 0.5 M HCl (15 min, RT), gelatinized (20 h, 75 °C), Ezee filtered and freeze-dried for 48 h. The weight of the resulting "bulk collagen" collected corresponds to the AG yield (AG refers to the protocol code used at ORAU[20]). If the sample was deemed sufficiently well preserved, that is, AG yield exceeded 10 mg, the "bulk collagen" was subsequently hydrolyzed in 10 mL of ultrapure water, and filtered using 30 kDa MWCO ultrafilters (Vivaspin, UK) until circa 1.5 mL of solution remained. The retentate was freeze-dried for 24 h, and the "purified collagen" weighed, thus giving the AF yield. With the weight of the original bone powder before treatment considered 100%, the yield for both AG and AF "collagen" could be calculated. Where no information for the AF yield is provided in Table 1, the sample could not be ultrafiltered as a result of low "collagen" preservation determined from the AG yield. From a radiocarbon perspective, an AG yield of <1 wt % was classified as very poor preservation, 1−3 wt % as poor preservation (but would be dated), 3−6 wt % as low preservation, and 6−8 wt % as good preservation. Archaeological samples with an AG "collagen" yield of >10 wt % would be seen as well preserved. For comparison, fresh bone would result in a "collagen" yield of approximately 22 wt %.[11]

RESULTS

Protein Extraction

The 44 LC−ESI−MS/MS runs overall allowed for the identification of 93 proteins (0.0% decoy FDR) and 29,735 spectra (0.00% decoy FDR). 36 out of 93 were collagenous proteins, eight were keratins, two were trypsins, and two belonged, respectively, to bacteria (RL7, *Mycobacterium bovis*) and yeast (ADH1, *Saccharomyces cerevisiae*). The intrinsic NCPs identified within the samples were therefore 45 in total. After protein BLAST (Basic Local Alignment Search Tool) checks to confirm endogeneity (i.e., to the species determined by ZooMS), we reduced this to a list of 36 NCPs (Table 2). Overall, 28% were blood/serum proteins, 58% were proteins found in bone tissue, 8% were extracellular proteins, and 6% were intracellular. The most common NCPs identified (based on total spectra counts) were biglycan (BGN), followed by albumin (ALB), pigment epithelium-derived factor (SER-PINF1), thrombospondin (THBS1), prothrombin (F2), chondroadherin (CHAD), nucleobindin-1 (NUCB1), SPARC, and fetuin-A (AHSG) (Table 2).

Comparison "Collagen Yield" and Number of NCPs

After refinement of the NCPs list per sample, the number of intrinsic NCPs obtained was compared with the derived "collagen yield" taken from ORAU analyses (specifically the "AG" yield as described in the Experimental Section, under "Collagen Yield") (Figure 2 and Table 3) and the total spectral counts matched with collagen α-1(I) (hereafter COL1A1) and α-2(I) (hereafter COL1A2) were also compared with the "% AG collagen yield" (Figure 3). Results showed that "collagen yield" and proteome variety were not correlated (Pearson's correlation p-value = 0.8833 and correlation coefficient = 0.0228). Specifically, several samples that generated a "collagen yield" of <1% contained up to 14 NCPs (TD1.B, Figure 2), whereas other samples with "collagen yields" of >7% contained less than four NCPs. Samples MR1.A and MR1.B generated the highest "AG collagen yield" in the dataset but contained three and zero NCPs respectively (Figure 2). Furthermore, "AG collagen yield" was not correlated with the total spectrum counts for COL1A1 and COL1A2 (Pearson's correlation p-value = 0.1965 and 0.4691 and correlation coefficient = 0.1985 and 0.1120, respectively), with the majority of the samples showing total spectrum numbers ranging between 234 and 380 for COL1A1 and between 149 and 277 for COL1A2 despite having different "collagen yield" values and with only two samples falling outside this range (KZ-58 and MR1.B). Pearson's correlation index calculated between the total spectrum count of the most abundant NCP (biglycan) and the "AG collagen yield" for each sample (e.g., considering each fraction generated from each bone as an individual sample) resulted in a nonsignificant correlation (p-value = 0.9606 and correlation coefficient = −0.0077) (Figure 4).

Comparisons between Fraction A and Fraction B

There were 15 samples for which it was possible to collect analysis material at two different stages of the radiocarbon dating process—namely after the second acid wash (fraction A) and after the overnight acid incubation (fraction B). For these samples, we analyzed the variability in the concentration of collagen and NCPs between those two fractions. While variations were present, no significant correlation was present (Figure 5). Despite some samples were characterized by similar amounts of COL1A1 but different amounts of NCPs in the

Table 2. List of all NCPs Identified in the 44 Ancient Samples ($n = 36$), Including Protein Function, Localization, and all Species Listed on Swiss-Prot That Have a 100% Amino Acid Sequence Match

protein ID	name	NCP contained in samples:	role	localization	species[a,b]
ALB	serum albumin	TD4.A–TD4.B–TD1.A–TD1.B–KZ-24–KZ-52–TD17.A–TD6.A–TD6.B–TD19.A–TD19.B–TD16.A–TD16.B–TD7.A–TD7.B–MAN1.A–MAN1.B–KZ-23–TD14.A–TD14.B–KZ-53–KZ-19–TD20.A–TD20.B–TD8.A–TD8.B–TD15.A–TD15.B	binding protein	serum	bovine, horse, human, cervine
APOA1	apolipoprotein A-I	TD4.B–TD1.A–TD1.B–KZ-19	transport of cholesterol	serum	bovine
APOA2	apolipoprotein A-II	TD1.A	stabilize high density lipoprotein	serum	bovine
F10	coagulation factor X	KZ-54–KZ-53	blood coagulation	serum	human
F7	coagulation factor VII	KZ-53	blood coagulation	serum	human
F9	coagulation factor IX	TD17.A	blood coagulation	serum	bovine
GC	vitamin D-binding protein	TD4.B–TD1.B	vitamin D transport	serum	bovine
PROS1	vitamin K-dependent protein S (fragment)	TD19.A–TD14.B	blood coagulation	serum	horse
SERPINC1	antithrombin-III	TD4.B–TD1.A–TD1.B	blood coagulation	serum	bovine
SERPIND1	heparin cofactor 2	TD19.A	blood coagulation	serum	horse
AHSG	α-2-HS-glycoprotein/fetuin-A	TD4.A–TD4.B–TD1.A–TD1.B–TD17.A–MAN1.B–KZ-53–KZ-19–TD8.A–TD8.B	skeletal mineralization	bone/serum	bovine, human
ALPL	alkaline phosphatase, tissue-nonspecific isozyme	TD4.B–TD19.A	skeletal mineralization	bone	bovine, horse
BGN	biglycan	TD4.A–TD4.B–TD1.A–TD1.B–MR8.B–TD5.A–TD5.B–KZ-54–KZ-24–KZ-52–TD17.A–TD17.B–TD6.A–TD6.B–TD19.A–TD19.B–TD16.A–TD16.B–TD7.A–TD7.B–MAN1.A–MAN1.B–KZ-23–TD14.A–TD14.B–KZ-53–KZ-19–TD20.A–TD20.B–TD8.A–TD8.B–TD15.A–TD15.B–MR1.A	collagen assembly	bone/ECM	bovine, horse, human, cervine
C3	complement C3	TD4.A–TD1.A–TD1.B–TD19.B–TD14.B	role in osteogenesis	bone/serum	bovine, horse
C9	complement component C9	TD19.A–TD19.B	role in osteogenesis	bone/serum	horse
CLEC11A	C-type lectin domain family 11 member A	TD19.A–TD19.B–TD7.A–TD14.B	promoting osteogenesis	bone	bovine, horse
CLEC3B	tetranectin	TD4.A–TD4.B–TD1.B–KZ-24–TD17.A–MAN1.B–KZ-23	bone mineralization	bone/serum	bovine, human, cervine
DPT	dermatopontin	KZ-53	collagen fibril formation	bone	human
F2	prothrombin	TD4.A–TD4.B–TD1.B–KZ-54–TD17.A–TD17.B–TD19.A–TD19.B–MAN1.A–MAN1.B–TD14.A–TD14.B–KZ-53–KZ-19–MR1.A	blood coagulation	bone/serum	bovine, horse, human
HTRA1	serine protease HTRA1B	TD1.A–TD1.B–TD19.A	osteogenesis regulation	bone	bovine, horse
IBSP	bone sialoprotein	TD6.B–TD16.A	binding mineral matrix	bone	horse
LUM	lumican	TD17.A–MAN1.A–MAN1.B–KZ-53	collagen binding	bone/ECM	bovine, human
NUCB1	nucleobindin-1	TD4.A–TD4.B–TD1.A–TD1.B–KZ-24–KZ-52–TD19.A–TD19.B–TD16.A–KZ-23–TD14.A–TD14.B–KZ-19–TD8.A	bone matrix maturation	bone	bovine, horse, cervine
OMD	osteomodulin	MAN1.A–MAN1.B–KZ-53	bone remodelling	bone	human
PANX3	pannexin-3	KZ-24	bone growth regulation	bone	cervine
POSTN	periostin	TD19.A–MAN1.B–TD14.A–TD14.B–KZ-53	bone remodelling	Bone	horse, human

Table 2. continued

protein ID	name	NCP contained in samples:	role	localization	species[a,b]
SERPINF1	pigment epithelium-derived factor	TD4.B–TD1.B–TD17.A–TD17.B–TD6.A–TD6.B–TD19.A–TD19.B–TD16.A–TD16.B–TD7.A–MAN1.A–MAN1.B–TD14.A–TD14.B–KZ-53–TD20.A–TD20.B–TD8.A–TD8.B- TD15.A–TD15.B	bone remodelling	bone/ECM	bovine, horse, human
SPARC	SPARC	TD4.A–TD4.B–TD1.B–TD6.A–TD6.B–TD19.A–TD19.B–TD16.A–TD16.B–TD7.A–TD14.A–TD14.B–TD20.A	calcium and collagen binding	bone	bovine, horse
SPP24	secreted phosphoprotein 24	TD17.A	bone turnover	bone/serum	bovine
THBS1	thrombospondin-1	TD4.A–TD4.B–TD1.A–TD1.B–KZ-24–KZ-52–TD19.A–TD19.B–TD7.A–MAN1.A–MAN1.B–KZ-23–TD14.A–TD14.B–KZ-53–KZ-19–TD8.A–TD8.B	bone homeostasis	bone	bovine, horse, human, cervine
VTN	vitronectin	KZ-54–TD19.B–KZ-53–MR1.A	bone mineralization	bone	horse, human
CHAD	chondroadherin	TD1.A–TD17.A–TD6.A–TD6.B–TD16.A–TD16.B–MAN1.A–MAN1.B–KZ-23–TD14.B–KZ-53–TD20.A–TD20.B–TD15.B	promoting chondrocyte growth	cartilage/extracellular	bovine, horse, human, cervine
DSP	desmoplakin	KZ-53	desmosomes assembly	intracellular	human
IGFALS	insulin-like growth factor-binding protein complex acid labile subunit	KZ-53	protein–protein interaction	extracellular	human
RDX	radixin	TD14.B	actin binding	intracellular	horse
S100A7	protein S100-A7	TD15.A	calcium binding	extracellular	horse

[a]Bovine = cattle and bison. [b]Cervine = red deer, fallow deer or elk.

two fractions (e.g., MAN1, TD16, TD1, and TD4), others contained different amounts of COL1A1 but similar amounts of NCPs (e.g., TD5). Furthermore, some samples were characterized by similar amounts of COL1A1 and NCPs (e.g., TD19, TD20) (Figure 5).

As there are differences in the degradation rate of collagen and post-translational protein modifications (PTMs), and there may be differences in the degradation process between individual samples, we compared the coverage values obtained from fraction A and fraction B for COL1A1 and COL1A2, as well as the three most abundant NCPs found in our dataset (BGN, ALB, and AHSG). To avoid biases due to differences in the annotation of the protein databases of different species, this evaluation was limited to one species only, *Bos taurus*. Results showed an absence of any specific trends, with some proteins showing deeper coverage in fraction A compared to fraction B, and vice-versa (Table 4).

Species Identification

All 29 bone samples were attempted to be morphologically identified prior to molecular experimentation (Table 1). Of these, five were identified to species-level (four as "human"; one as "red deer"), one to subfamily ("Bovinae"), seven to clade ("ungulate"), 13 to class ("mammal"), and one to order ("Artiodactyla"). Two samples could not be morphologically identified. ZooMS (via MALDI-TOF MS) was able to refine the taxonomic identifications of all collagen-containing samples (22 of 29) to bovine (cattle or bison) ($n = 6$), horse ($n = 9$), human ($n = 2$), and cervine (e.g., red deer, fallow deer or elk) ($n = 5$) (Figure 2–4, full-shaped points). Finally, LC–ESI–MS/MS data was used to further refine the taxonomic classifications, verifying and/or generating species-level identifications for 25 of the 29 samples, with the remaining suspected as Bovidae/Cervidae with this identification only limited by the available proteomes on Swiss-Prot, which does not currently contain proteomic sequences for cervine (Figure 2–4, empty-shaped points).

Protein Diagenesis

Semitryptic searches that were performed to evaluate the extent of diagenesis (Table S2) showed a consistent percentage of total spectrum counts for COL1A1 and COL1A2 peptides regardless of whether they were from high- or low-protein yield samples; the ratio of spectral counts identified using standard tryptic searches to the ones found using semitryptic searches was on average 46.6% and 46.7% for COL1A1 and COL1A2, respectively, in the 10 lowest protein yield samples, and 44.9% and 46.9% in the 10 highest protein yield samples, respectively.

■ DISCUSSION

In the following, we discuss our findings on the endogeneity of the dominant proteinaceous biomolecules within archaeological samples typically considered by radiocarbon and stable isotope specialists to be poorly preserved. Following our main aims, we focused on variability between samples and fractions, correlations between proteomic results and radiocarbon "collagen yield" and species identification. First, we examine whether our decision on which waste fraction from the radiocarbon dating pretreatment to sample influences our analysis results. Second, we discuss the influence of bone diagenesis. This includes a comparison between the number of identified NCPs and sample age and between radiocarbon "collagen yield" (AG) and protein coverage. We subsequently extend this to focus on nonspecific trypsin cleavages caused by
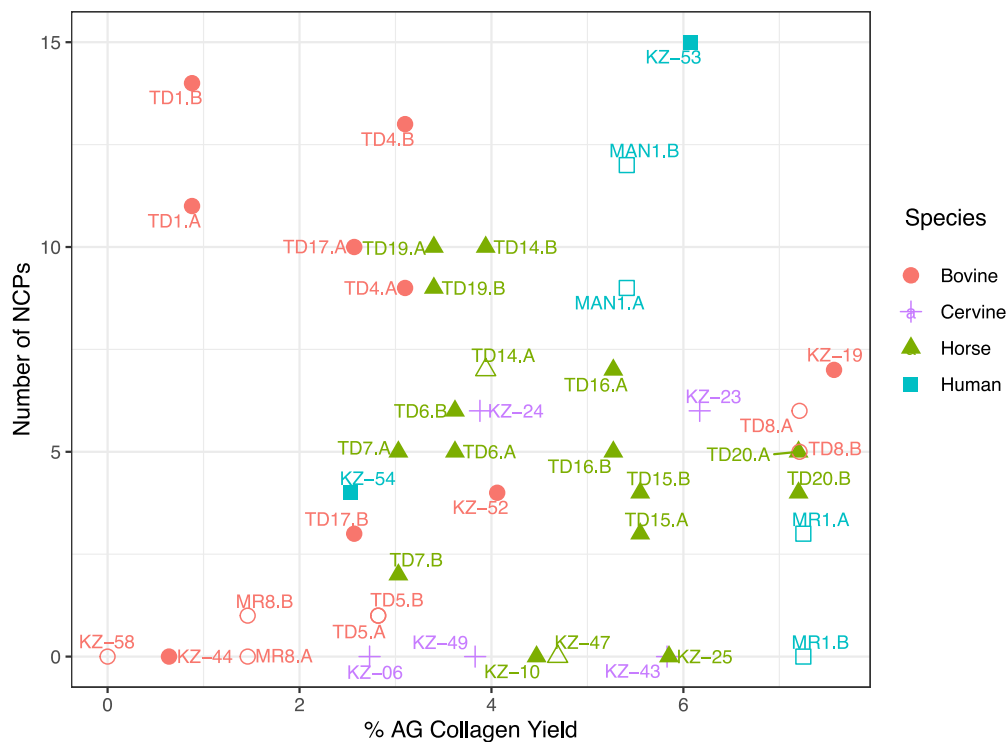
Figure 2. Scatterplot for the percentage of "AG Collagen yield" (*X* axis) and for the number of NCPs (*Y* axis) identified in the dataset. Different colors and symbols (legend) represent different species. Species identifications achieved using both ZooMS and LC−ESI−MS/MS proteomics data are indicated with full shapes, whereas species identifications achieved uniquely with LC−ESI−MS/MS proteomics (and where ZooMS analyses failed) are indicated with empty shapes.

diagenesis. Third, we compare the success and nature of species identification results obtained through traditional (i.e., morphological) means, ZooMS, and proteomic analysis. Lastly, we take a closer look to assess how our results compare with other findings from the literature.

Proteomic Differences between Radiocarbon Dating Fractions

In this study, we compared the fractions obtained from the second HCl prewash step (fraction A) and the overnight HCl incubation (fraction B) to provide a comparison between the two processing stages (Table 1).

While we did not observe any significant correlation between the concentration of collagen or NCPs and the fraction sampled, we did notice that, in general, greater variations between the two fractions can be observed for NCPs compared to COL1A1 abundance. In particular, NCPs were more abundant in fraction A than in fraction B in seven cases and were identical in one case. These results may be related to differences in the degradation rate of collagen and PTMs and in the diagenetic processes that could have affected the samples, with more fragmented proteins tending to be released during the second prewashing (fraction A) and with more intact proteins being released within fraction B. However, when comparing coverage values (restricted to *Bos taurus*) for COL1A1, COL1A2, and for three of the most abundant PTMs found in the dataset (BGN, ALB, and AHSG), no specific trends could be identified: some proteins showed a deeper coverage in fraction A and lower coverage in fraction B and vice-versa (Table 4).

We also did not observe any consistent change in the degradation of the proteome between specific fractions, as showed by the semitryptic searches (Table S2). For example,

in some cases, fraction A had an increased percentage of collagen semitryptic peptides, and in other cases, fraction B had a higher amount of those. This may be due to the fact that bone demineralization was not complete after the second HCl treatment (fraction A), retaining sufficient material for proteome recovery after the overnight incubation (fraction B). By contrast, a previous study on modern bones showed that a demineralization length of 6 h allowed for a better proteome recovery than prolonged lengths (24 and 48 h).[43] It is worth emphasizing here that the radiocarbon samples had already completed a first HCl treatment for 2 h, followed by discarding of the soluble fraction of proteins, prior to the HCl treatment from which fraction A had been collected. We believe that this combined time has been enough to demineralize the ancient samples sufficiently to extract a high number of NCPs and that the subsequent overnight incubation step did not significantly improve the overall extraction of the proteins embedded in the mineral matrix, nor did it significantly increase the protein damage induced by the interaction with the acid.

Looking at the common contaminants usually found in bone samples such as keratins, we found that seven out of 15 samples contained keratins only in fraction A, six samples contained keratins in both fractions and only two samples contained keratins exclusive to fraction B (Table S3). These results show that, although the pretreatment step can remove some of the common modern contaminants known to affect the dating results, further steps should be carried out to ensure removal of all of them, otherwise contamination with modern carbon would still be expected in the gelatinized fraction.

Table 3. "AG Collagen Yield" (See Experimental Section, under Collagen Yield), Number of Total Spectrum Count for Collagen α-1(I) and Collagen α-2(I) Chains, Number of NCPs, and Percentage of Coverage for Collagen α-1(I) and Collagen α-2(I), Biglycan (BGN) and Albumin (ALB)

sample	AG collagen yield	COL1A1	COL1A2	NCPs	coverage COL1A1 (%)	coverage COL1A2 (%)	coverage BGN (%)	coverage ALB (%)
KZ-58	0	13	13	0	11	11		
TD4.A	3.1	284	194	9	47	45	11	12
TD4.B	3.1	267	183	13	51	51	15	16
KZ-44	0.64	357	277	0	50	54		
TD1.A	0.88	296	234	11	51	55	18	21
TD1.B	0.88	307	223	14	49	50	15	21
MR8.A	1.46	241	201	0	47	48		
MR8.B	1.46	306	216	1	45	47	7	
KZ-06	1.6	355	227	0	51	46		
TD5.A	2.22	259	149	1	50	46	7	
TD5.B	2.22	316	192	1	44	47	7	
KZ-54	2.53	295	152	4	47	35		
KZ-49	2.68	376	211	0	50	39	11	5
KZ-24	2.76	334	179	6	49	53	11	8
KZ-52	2.76	347	261	4	51	53	16	9
TD17.A	3.14	340	261	10	51	52	13	
TD17.B	3.14	315	226	3	70	72		
KZ-10	3.17	294	227	0	47	43		
KZ-43	3.24	343	253	0	77	80	14	9
TD6.A	3.43	313	247	5	61	80	15	5
TD6.B	3.43	312	277	6	83	77	29	28
TD19.A	3.5	295	247	10	77	77	18	21
TD19.B	3.5	313	255	9	72	77	19	8
TD16.A	3.55	243	219	7	69	72	19	6
TD16.B	3.55	265	218	5	75	81	15	6
TD7.A	3.68	254	231	5	75	74	11	4
TD7.B	3.68	297	251	2	52	56	14	8
MAN1.A	4.09	359	256	9	49	58	18	8
MAN1.B	4.09	368	247	12	72	79		
KZ-25	4.17	349	258	0	51	37	14	5
KZ-23	4.47	328	190	6	73	63		
KZ-47	4.59	234	176	0	76	80	27	27
TD14.A	4.85	308	260	7	75	78	29	27
TD14.B	4.85	274	266	10	51	56	23	8
KZ-53	4.95	380	243	15	52	54	11	8
KZ-19	5.16	355	233	7	75	76	14	6
TD20.A	5.28	312	230	5	79	69	14	6
TD20.B	5.28	295	201	4	49	58	7	5
TD8.A	5.73	263	218	6	47	51	11	5
TD8.B	5.73	324	222	5	47	42	7	
TD15.A	6.19	350	196	3	78	80	7	4
TD15.B	6.19	294	275	4	74	82	11	4
MR1.A	6.43	282	191	3	51	52	7	
MR1.B	6.43	134	93	0	41	41		

NCP Presence versus Sample Age

The samples that contained ten or more NCP matches (TD1, TD4, TD14, TD17, TD19, MAN1, and KZ-53) range in age from late Holocene samples (KZ-53 and MAN1) to late Middle Palaeolithic and early Upper Palaeolithic (TD1, 4, 14, 17, and 19), yielding no clear correlation between the age of the samples and the number of NCPs. For example, there are some samples richer in NCPs that are older than others with fewer NCPs. Sample KZ-54 dated to the late Holocene had four NCPs (fraction B), whereas TD4, one of the oldest specimens, had nine and 13 NCPs extracted from the prewash (fraction A) and from the acid incubation fraction (fraction B), respectively. Clearly the depositional environment would have

played a major role in the survival of the NCPs in the specimens. For example, although samples MR1 and MR8 were excavated from the same cave, depth, and level, they were dated to the late Holocene and late Middle Palaeolithic to early Upper Palaeolithic periods, respectively, and they contained three and one NCP only, respectively, in both the analyzed fractions (A plus B). This suggests that the taphonomic processes that affected the bone proteome survival were more likely to be related to environmental factors than to aging phenomena.

"Collagen Yield" and Protein Coverage

We did not observe any obvious trends between a sample's "collagen yield" (as determined by the radiocarbon dating
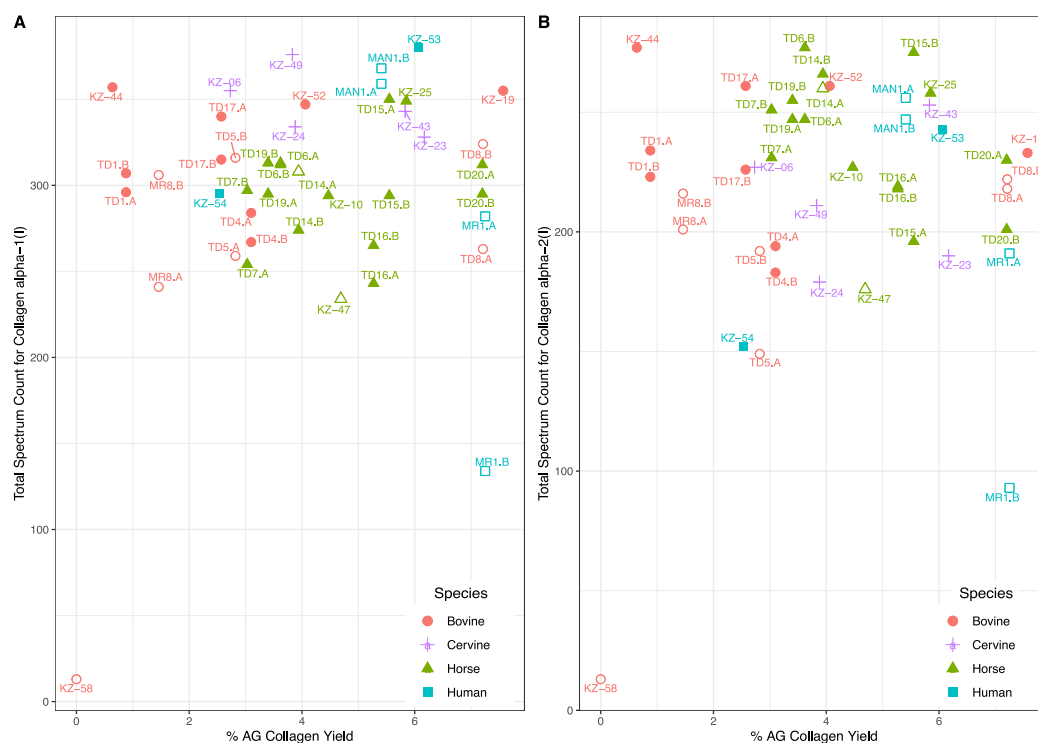
Figure 3. Scatterplot for the percentage of "AG collagen yield" (X axis) and for the total spectrum counts for (A) COL1A1 and B) COL1A2 (Y axis) identified in the dataset. Different colors and symbols (legend) represent different species. Species identifications achieved using both ZooMS and LC−ESI−MS/MS proteomics data are indicated with full shapes, whereas species identifications achieved uniquely with LC−ESI−MS/MS proteomics (and where ZooMS analyses failed) are indicated with empty shapes.
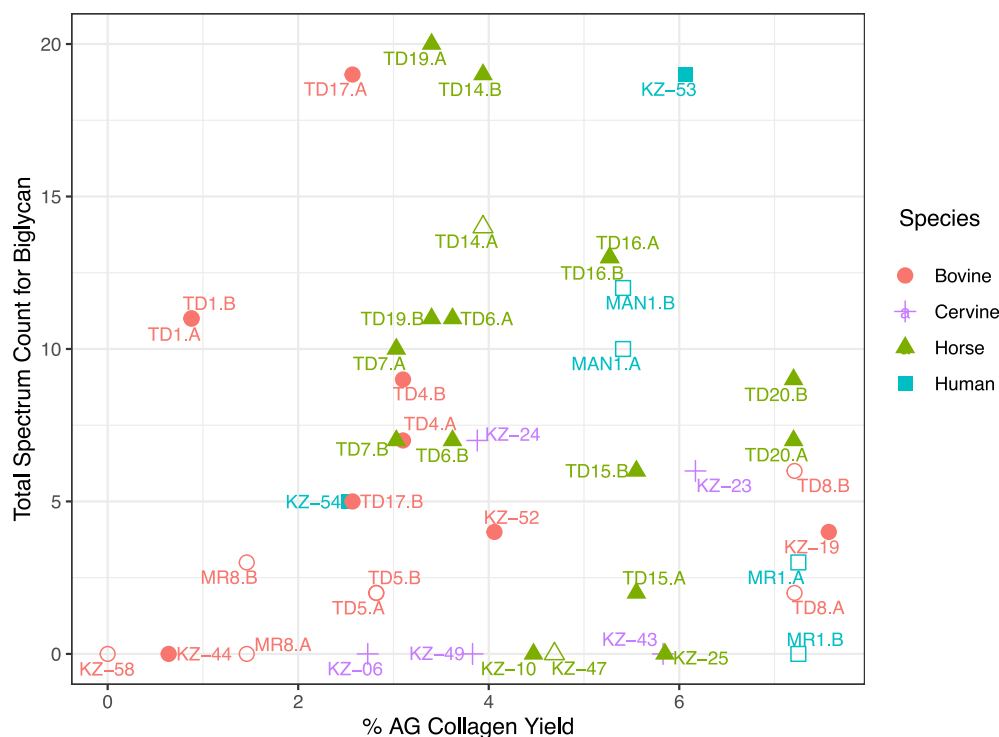

Figure 4. Scatterplot for the percentage of "AG Collagen Yield" (X axis) and for the total spectrum counts for biglycan (Y axis) identified in the dataset. Different colors and symbols (legend) represent different species. Species identifications achieved using both ZooMS and LC−ESI−MS/ MS proteomic data are indicated with full shapes, whereas species identifications achieved uniquely with LC−ESI−MS/MS proteomics (and where ZooMS analyses failed) are indicated with empty shapes.

process) and the proteome complexity. We did identify specific globular serum proteins (such as albumin and fetuin-A)

together with collagen-binding proteins (such as biglycan) in the majority of the samples, despite the large range in "collagen
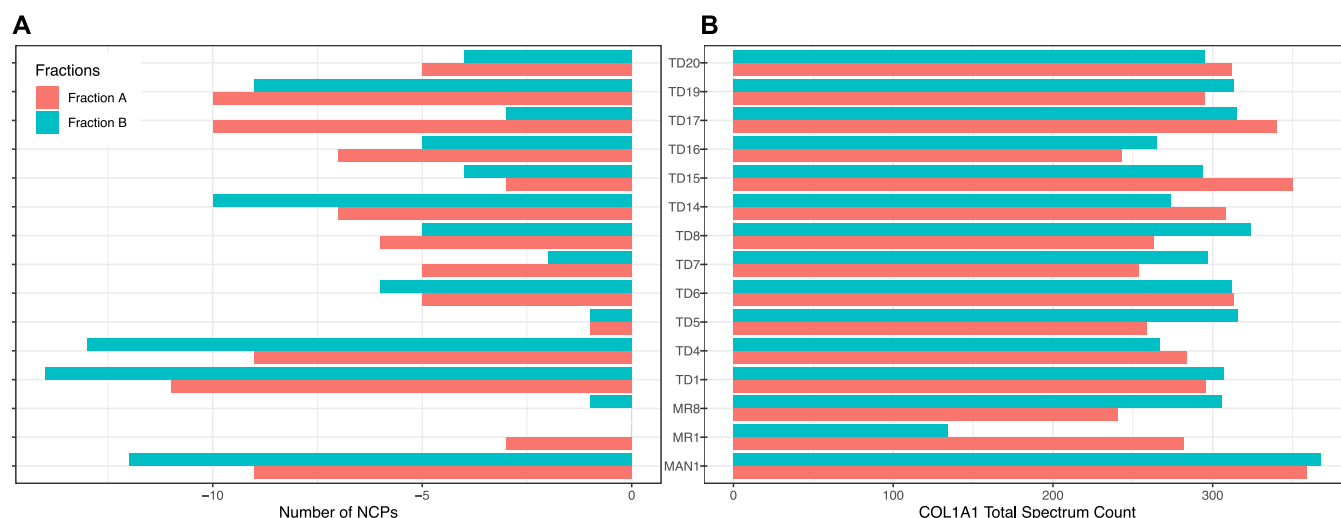
Figure 5. Bar plot representing (A) number of NCPs and (B) COL1A1 total spectrum count obtained from fraction A (red) and fraction B (blue) of samples subjected to a prewashing step with HCl. Sample names are indicated to the side of the two bar plots.

Table 4. Percentage Coverage for Collagen α-1(I) Chain, Collagen α-2(I) Chain, Biglycan, Serum Albumin and Fetuin-A for Fraction A and B for Bovine Samples MR8, TD1, TD17, TD4, and TD5

	MR8.A (%)	MR8.B(%)	TD1.A(%)	TD1.B (%)	TD17.A (%)	TD17.B (%)	TD4.A (%)	TD4.B(%)	TD5.A (%)	TD5.B (%)
COL1A1	47	45	51	49	51	51	47	51	47	50
COL1A2	48	47	55	50	53	52	45	51	42	46
BGN			18	15	16	13	11	15	7	7
ALB			21	21	9		12	16		
AHSG			12	14	10		16	16		

yield" that the samples demonstrated (0.0−7.57%; AG collagen yield). To further investigate this lack of correlation, we also explored the percentage coverage for COL1A1, COL1A2, BGN, and ALB (Table 3). For this, we excluded samples identified as a horse ($n = 17$) due to an incompleteness of the database that would not allow a reliable evaluation of the percentage coverage of each protein. The percentage protein coverage ranged between 41 and 52% for COL1A1 and between 37 and 58% for COL1A2 and did not follow any specific trends related with the "collagen yield" of the samples. For example, the highest "collagen" yielding sample, MR1.B, had 41% coverage for both COL1A1 and COL1A2, whereas one of the lowest "collagen" yielding samples, TD1.A, had 51 and 55% coverage, respectively (Table 3). The only exception was sample KZ-58, which yielded no collagen and poor coverage for both COL1A1 and COL1A2 (11% for both chains). The same trend was observed for the most abundant PTMs found in the dataset, namely BGN and ALB, which did not show any correlation between increasing "collagen yield" between samples and percentage coverage.

Protein Diagenesis

Results on the percentage of semitryptic peptides found in our samples showed a relatively high level of those (average ~45%) in comparison with the amounts usually obtained when operating in optimal conditions (e.g., extracting proteins from fresh and modern tissues); the percentage of semitryptic peptides was very low, from less than 3% for soft tissues[44] to around 15% for hard tissues, where a demineralization step similar to the one used in this study is required to allow for protein extraction.[43] The frequency of diagenetically altered peptides can increase during both preparation and digestion of samples, depending on the protocol used for the extraction;

however, in this case, results were notably higher than the percentages expected from modern samples. We also made a comparison among the percentage of semitryptic peptides and the collagen yield obtained from our samples and we noted that, overall, collagenous proteins accumulate damage over time in archaeological timeframes, regardless of the total amount of collagen that can be extracted from the samples. For BGN, we found that percentages for semitryptic peptides averaged 47.4% for the ten lowest "collagen yield" samples (as defined by radiocarbon) and 48.5% for the ten highest "collagen yield" samples. This result suggests that globular proteins are subjected to a very similar decay rate to that of collagen and shows that the damage of NCPs is not directly related with the amount of "collagen" extracted during the radiocarbon dating pretreatment.

ZooMS and Cross-Species Proteomics

Focusing specifically on low "collagen yield" samples (as defined by the radiocarbon dating analysis), we had three samples classified as having a yield of less than 1%, namely samples KZ-58, KZ-44, and TD1. KZ-58 was a morphologically unidentified mammal bone from Upper Paleolithic contexts of the Kozarnika site (Bulgaria) whose ZooMS analysis also failed in identifying its species. Although shotgun proteomic analysis revealed that the specimen of interest could be attributed to a bovine, a further search against a local database derived from protein BLAST searches of the cattle COL1A1 and COL1A2 sequences confirmed a greater match to a cervid sequence (with notable matches to peptide sequences GETGPSGPAGPTGAR, GAPGAVGAPGPA-GANGDR, and TGQPGAVGPAGIR differentiating them from cattle); interestingly, no NCPs were identified in this sample, consistent with its relatively poor molecular survival
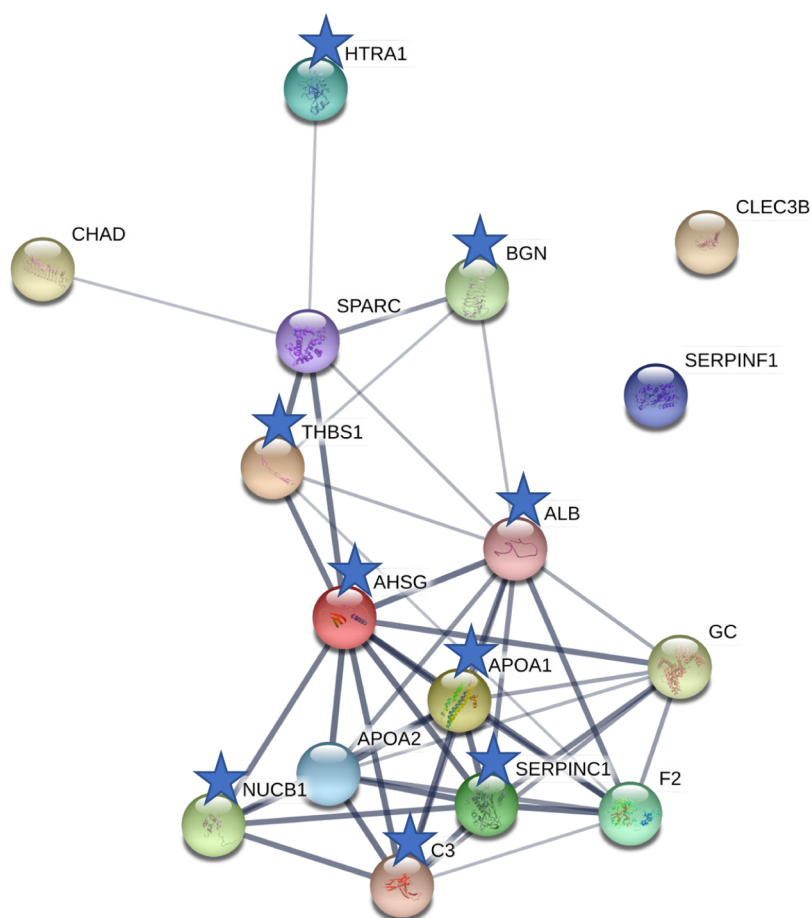
Figure 6. STRING association network of the NCPs extracted from TD1.A and TD1.B. The line thickness indicates the strength of data support (edge confidence). Proteins marked with the star symbol were identified in both sample fractions A and B (second prewash and final overnight HCl incubation).

and the notion that collagen typically survives longer. Likewise, sample KZ-44 (collected from Kozarnika and considered Upper Paleolithic), which was visually attributed to a large ungulate identified by ZooMS as being *Bos* spp., also yielded no NCPs, but it did yield a higher abundance of collagenous spectra compared to KZ-58 (357 versus 13 spectra for COL1A1 and 277 versus 13 spectra for COL1A2). TD1, an unidentified mammal bone from Upper Paleolithic levels of Temnata, which was identified by ZooMS as *Bos* spp., had a lower abundance of collagenous spectra (both fractions) when compared with sample KZ-44 (COL1A1 296 and 307 spectra, for TD1.A and TD1.B, respectively, versus 357, and COL1A2 223 and 201 spectra versus 277) but nonetheless yielded >10 NCPs (Figure 6).

The limitation of the LC−MS/MS identification in comparison with ZooMS is the incompleteness of the available proteomic databases for some species of animal, which, for example, does not allow for the distinction of bovine from cervine samples (Table 1). Conversely, advantages for the use of LC−MS/MS proteomic analyses in combination with ZooMS approaches include the possibility to look at peptides of NCPs for the identification of specimens characterized by a poor collagen fingerprint spectrum. Among all NCPs identified in this work, albumin, biglycan, thrombospondin-1, and chondroadherin were the ones that were identified in each of the four species present, and pigment epithelium-derived factor and prothrombin were identified in three out of four species

(specifically in bovine, horse, and human) (Table 2). Fetuin-A, a protein normally identified in ancient bones, was not found in any horse or cervine samples. We believe that this result is potentially due to the lack of completeness of the proteomic databases for these species rather than to the decay or failure in the extraction of this specific protein in these animal species. Because of the great potential that fetuin-A has to extract phylogenetic information from samples, we suggest the creation of an ad-hoc database with fetuin-A sequences for the species of interest in order to allow for the identification and matching of its peptides and finally its use to conduct phylogenetic and species identification studies. Interestingly, fetuin-A was successfully identified in both of the two fractions in one of the lowest "collagen yield" samples (TD1, whose "collagen yield" was between the second-lowest found in this work at 0.88%) that would normally have been discarded by ORAU for subsequent radiocarbon dating due to the scarce reliability that the measurements would have in these cases. Further research may help to better evaluate whether a sample could provide a reliable date, despite the low radiocarbon "collagen yield".

Influence of Protein Extraction Protocols from Ancient Bone

When comparing our findings with previous work[34] where acid-insoluble pellets were treated with guanidine hydro-chloride (GuHCl) after the overnight incubation step in 0.6 M HCl and only this fraction (comparable to fraction B in this

study) was analyzed, we found that the variety of NCPs found in our batch was smaller (maximum number of 15 NCPs, versus 30 NCPs found in the previous work), despite the totals over the dataset being similar, with 37 and 44 NCPs, respectively. This difference in protein number further supports the suggestion that incubation in GuHCl rather than HCl is a valuable method to increase the number of identified NCPs in bone samples, as has also been shown in a study conducted on ancient bovid teeth and mandible bones.[36] The most commonly identified NCPs in this study were the same as those found by Wadsworth and Buckley[34] despite the fact that in this study, there was no incubation step in GuHCl. The only two exceptions to this were the NCPs thrombospondin and SPARC (commonly found in this study but not in ref [34]) and for lumican (mentioned in ref [34] but found only in a limited number of samples here).

Observing the methods achieved in previous work,[19] four bovid specimens (dated from ∼4 to 130 Ka) were treated with a similar protocol to the one used here, omitting the pretreatment using solvents but including a prewash step with 0.6 M HCl for 2 h at R/T (similar to fraction A) prior to incubation with the same acid overnight at R/T (similar to fraction B), with the two acid fractions having been pooled together, "RC sol-fraction". These results were comparable to those of this study. In particular, from four to ten NCPs were previously observed in the "RC sol-fraction", and fetuin-A, PEDF, ALB, biglycan, lumican, complement C3, decorin, and prothrombin were the most commonly identified NCPs, despite not being found in all samples. Despite the fact that the two acid fractions were not combined in our work, the comparability of the results obtained with other proteomic analyses on ancient bones[19] suggests that either of the acid soluble fractions generated during the processing of the samples for radiocarbon dating contain a substantial variety of NCPs that can be used for phylogenetic purposes and for cross-proteomics analyses, as well as potentially for isotopic and radiocarbon studies.

CONCLUSIONS

Overall, our results show that the indication provided by the "collagen yield" of archaeological samples (as defined in radiocarbon and stable isotope studies) should be used with caution, in that what may be considered "poor collagen" specimens for isotopic purposes may not necessarily be so for yielding proteomic information. Furthermore, our results support the previous studies highlighting that even the fractions that are typically discarded during the collagen extraction process can yield useful proteomes, with both the prewash and the acid incubation fractions containing several NCPs that can be successfully used to determine species identity. Moreover, the proteins contained in the acid fractions may be adequate to conduct isotopic studies and radiocarbon dating of the specimens; in fact, results showed that the total number of spectra found for either collagen α-1, collagen α-2, and for NCPs can be sufficient to conduct these types of studies despite the poor "collagen yield" calculated from the samples and the complete lack of correlation between these two variables. We have not found a clear correlation between proteome variety and age of the specimens (which stands in contrast with other findings from other datasets) but rather that the depositional environment played a more important role in the survival of specific proteins over any aging phenomena. We also showed here that both "fraction A" and

"fraction B", originated during the collagen extraction methodology, can contain a high number of NCPs. Additionally, the overall level of protein decay in the two fractions is comparable and common contaminants, such as keratins, are less abundant in the second fraction than in the first one. Finally, we show that LC−MS/MS proteomic analysis can be valuable in identifying samples that fail species identification through ZooMS collagen peptide mass fingerprinting.

■ ASSOCIATED CONTENT

ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jproteome.0c01014.

> Supporting Table S1: Accession names found on PRIDE and associated sample names found in the manuscript; Supporting Table S2: Full tryptic peptides (total spectrum count) and semitryptic peptides (total spectrum count) for COL1A1, COL1A2, and BGN for the ten lowest and from the ten highest collagen yield samples, percentages for the amount of semitryptic peptides and average for the ten lowest and ten highest collagen yield samples; Supporting Table S3: Keratins found in fraction A and fraction B (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Noemi Procopio − *Forensic Science Research Group, Faculty of Health and Life Sciences, Northumbria University, Northumbria University Newcastle, Newcastle Upon Tyne NE1 8ST, U.K.; Manchester Institute of Biotechnology, The University of Manchester, Manchester M1 7DN, U.K.;* ⓘ orcid.org/0000-0002-7461-7586; Phone: +44 (0)191 227 3468; Email: noemi.procopio@northumbria.ac.uk

Michael Buckley − *Manchester Institute of Biotechnology, The University of Manchester, Manchester M1 7DN, U.K.; Interdisciplinary Centre for Ancient Life, School of Natural Sciences, University of Manchester, Manchester, U.K.;* ⓘ orcid.org/0000-0002-4166-8213; Phone: +44(0)161 306 5175; Email: m.buckley@manchester.ac.uk

Authors

Rachel J.A. Hopkins − *Department of Anthropology, University of New Mexico, Albuquerque, New Mexico 87131-0001, United States; Research Laboratory for Archaeology and the History of Art (RLAHA), School of Archaeology, University of Oxford, Oxford OX1 3TG, U.K.*

Virginia L. Harvey − *Manchester Institute of Biotechnology, The University of Manchester, Manchester M1 7DN, U.K.; Interdisciplinary Centre for Ancient Life, School of Natural Sciences, University of Manchester, Manchester, U.K.*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jproteome.0c01014

Notes

ACKNOWLEDGMENTS

REFERENCES

(1) Ambrose, S. H. Preparation and characterization of bone and tooth collagen for isotopic analysis. *J. Archaeol. Sci.* **1990**, *17*, 431−451.

(2) Katzenberg, M. A.. Stable isotope analysis: a tool for studying past diet, demography, and life history. *Biological Anthropology of the Human Skeleton*; Wiley, 2008; pp 411−441.

(3) Bearhop, S.; Adams, C. E.; Waldron, S.; Fuller, R. A.; MacLeod, H. Determining trophic niche width: a novel approach using stable isotope analysis. *J. Anim. Ecol.* **2004**, *73*, 1007−1012.

(4) Fernandes, R.; Jaouen, K. Isotopes in archaeology. *Archaeol. Anthropol. Sci.* **2017**, *9*, 1305−1306.

(5) Schoeninger, M. J.; DeNiro, M. J.; Tauber, H. Stable nitrogen isotope ratios of bone collagen reflect marine and terrestrial components of prehistoric human diet. *Science* **1983**, *220*, 1381−1383.

(6) Turner Tomaszewicz, C. N.; Seminoff, J. A.; Ramirez, M. D.; Kurle, C. M. Effects of demineralization on the stable isotope analysis of bone samples. *Rapid Commun. Mass Spectrom.* **2015**, *29*, 1879−1888.

(7) Clarke, B. Normal bone anatomy and physiology. *Clin. J. Am. Soc. Nephrol.* **2008**, *3*, S131−S139.

(8) Longin, R. New method of collagen extraction for radiocarbon dating. *Nature* **1971**, *230*, 241−242.

(9) Brown, T. A.; Nelson, D. E.; Vogel, J. S.; Southon, J. R. Improved collagen extraction by modified Longin method. *Radiocarbon* **1988**, *30*, 171−177.

(10) Ramsey, C. B.; Higham, T.; Bowles, A.; Hedges, R. Improvements to the pretreatment of bone at Oxford. *Radiocarbon* **2004**, *46*, 155−163.

(11) van Klinken, G. J. Bone Collagen Quality Indicators for Palaeodietary and Radiocarbon Measurements. *J. Archaeol. Sci.* **1999**, *26*, 687−695.

(12) Maspero, F.; Sala, S.; Fedi, M. E.; Martini, M.; Papagni, A. A new procedure for extraction of collagen from modern and archaeological bones for 14C dating. *Anal. Bioanal. Chem.* **2011**, *401*, 2019−2023.

(13) Sealy, J.; Johnson, M.; Richards, M.; Nehlich, O. Comparison of two methods of extracting bone collagen for stable carbon and nitrogen isotope analysis: comparing whole bone demineralization with gelatinization and ultrafiltration. *J. Archaeol. Sci.* **2014**, *47*, 64−69.

(14) Semal, P.; Orban, R. Collagen Extraction from Recent and Fossil Bones: Quantitative and Qualitative Aspects. *J. Archaeol. Sci.* **1995**, *22*, 463−467.

(15) Liden, K.; Takahashi, C.; Nelson, D. E. The Effects of Lipids in Stable Carbon Isotope Analysis and the Effects of NaOH Treatment on the Composition of Extracted Bone Collagen. *J. Archaeol. Sci.* **1995**, *22*, 321−326.

(16) Arslanov, K. A.; Svezhentsev, Y. S. An Improved Method for Radiocarbon Dating Fossil Bones. *Radiocarbon* **1993**, *35*, 387−391.

(17) Caputo, I.; Lepretti, M.; Scarabino, C.; Esposito, C.; Proto, A. An acetic acid-based extraction method to obtain high quality collagen from archeological bone remains. *Anal. Biochem.* **2012**, *421*, 92−96.

(18) Cersoy, S.; Zazzo, A.; Lebon, M.; Rofes, J.; Zirah, S. Collagen Extraction and Stable Isotope Analysis of Small Vertebrate Bones: A Comparative Approach. *Radiocarbon* **2017**, *59*, 679−694.

(19) Wadsworth, C.; Buckley, M. Characterization of Proteomes Extracted through Collagen-based Stable Isotope and Radiocarbon Dating Methods. *J. Proteome Res.* **2017**, *17*, 429−439.

(20) Brock, F.; Higham, T.; Ditchfield, P.; Ramsey, C. B. Current pretreatment methods for AMS radiocarbon dating at the Oxford Radiocarbon Accelerator Unit (ORAU). *Radiocarbon* **2010**, *52*, 103−112.

(21) Brock, F.; Higham, T.; Bronk Ramsey, C. Radiocarbon dating bone samples recovered from gravel sites. *EH Res. Dept. Rep. Ser.* **2007**, *30*, 2007.

(22) Lebon, M.; Reiche, I.; Gallet, X.; Bellot-Gurlet, L.; Zazzo, A. Rapid quantification of bone collagen content by ATR-FTIR spectroscopy. *Radiocarbon* **2016**, *58*, 131.

(23) Le Meillour, L.; Zazzo, A.; Lesur, J.; et al. Identification of degraded bone and tooth splinters from arid environments using palaeoproteomics. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **2018**, *511*, 472−482.

(24) Bouchard, G. P.; Mentzer, S. M.; Riel-Salvatore, J.; et al. Portable FTIR for on-site screening of archaeological bone intended for ZooMS collagen fingerprint analysis. *J. Archaeol. Sci. Reports* **2019**, *26*, 101862.

(25) Pal Chowdhury, M.; Wogelius, R.; Manning, P. L.; Metz, L.; Slimak, L.; Buckley, M. Collagen deamidation in archaeological bone as an assessment for relative decay rates. *Archaeometry* **2019**, *61*, 1382−1398.

(26) Kontopoulos, I.; Penkman, K.; Mullin, V. E.; et al. Screening archaeological bone for palaeogenetic and palaeoproteomic studies. *PLoS One* **2020**, *15*, No. e0235146.

(27) Harvey, V. L.; Egerton, V. M.; Chamberlain, A. T.; Manning, P. L.; Buckley, M. Collagen Fingerprinting: A New Screening Technique for Radiocarbon Dating Ancient Bone. *PLoS One* **2016**, *11*, No. e0150650.

(28) Buckley, M.; Collins, M.; Thomas-Oates, J.; Wilson, J. C. Species identification by analysis of bone collagen using matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **2009**, *23*, 3843−3854.

(29) Collins, M.; Buckley, M.; Grundy, H. H.; Thomas-Oates, J.; Wilson, J.; van Doorn, N. ZooMS: the collagen barcode and fingerprints. *Spectrosc. Eur.* **2010**, *22*, 6.

(30) Brown, S.; Higham, T.; Slon, V.; et al. Identification of a new hominin bone from Denisova Cave, Siberia using collagen fingerprinting and mitochondrial DNA analysis. *Sci. Rep.* **2016**, *6*, 23559.

(31) Devièse, T.; Karavanić, I.; Comeskey, D.; et al. Direct dating of Neanderthal remains from the site of Vindija Cave and implications for the Middle to Upper Paleolithic transition. *Proc. Natl. Acad. Sci. U.S.A.* **2017**, *114*, 10606−10611.

(32) Buckley, M.; Whitcher Kansa, S.; Howard, S.; Campbell, S.; Thomas-Oates, J.; Collins, M. Distinguishing between archaeological

sheep and goat bones using a single collagen peptide. *J. Archaeol. Sci.* **2010**, *37*, 13−20.

(33) Brock, F.; Wood, R.; Higham, T. F. G.; Ditchfield, P.; Bayliss, A.; Ramsey, C. B. Reliability of nitrogen content (% N) and carbon: nitrogen atomic ratios (C: N) as indicators of collagen preservation suitable for radiocarbon dating. *Radiocarbon* **2012**, *54*, 879−886.

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

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

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

(37) Zhang, Y.; Fonslow, B. R.; Shan, B.; Baek, M.-C.; Yates, J. R., III Protein analysis by shotgun/bottom-up proteomics. *Chem. Rev.* **2013**, *113*, 2343−2394.

(38) Buckley, M.; Wadsworth, C. Proteome degradation in ancient bone: diagenesis and phylogenetic potential. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **2014**, *416*, 69−79.

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

(40) Procopio, N.; Chamberlain, A. T.; Buckley, M. Intra- and Interskeletal Proteome Variations in Fresh and Buried Bones. *J. Proteome Res.* **2017**, *16*, 2016−2029.

(41) Cersoy, S.; Zirah, S.; Marie, A.; Zazzo, A. Toward a versatile protocol for radiocarbon and proteomics analysis of ancient collagen. *J. Archaeol. Sci.* **2019**, *101*, 1−10.

(42) Buckley, M.; Harvey, V. L.; Chamberlain, A. T. Species identification and decay assessment of Late Pleistocene fragmentary vertebrate remains from Pin Hole Cave (Creswell Crags, UK) using collagen fingerprinting. *Boreas* **2017**, *46* (3), 402−411.

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

(44) Olsen, J. V.; Ong, S.-E.; Mann, M. Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Mol. Cell. Proteomics* **2004**, *3*, 608−614.