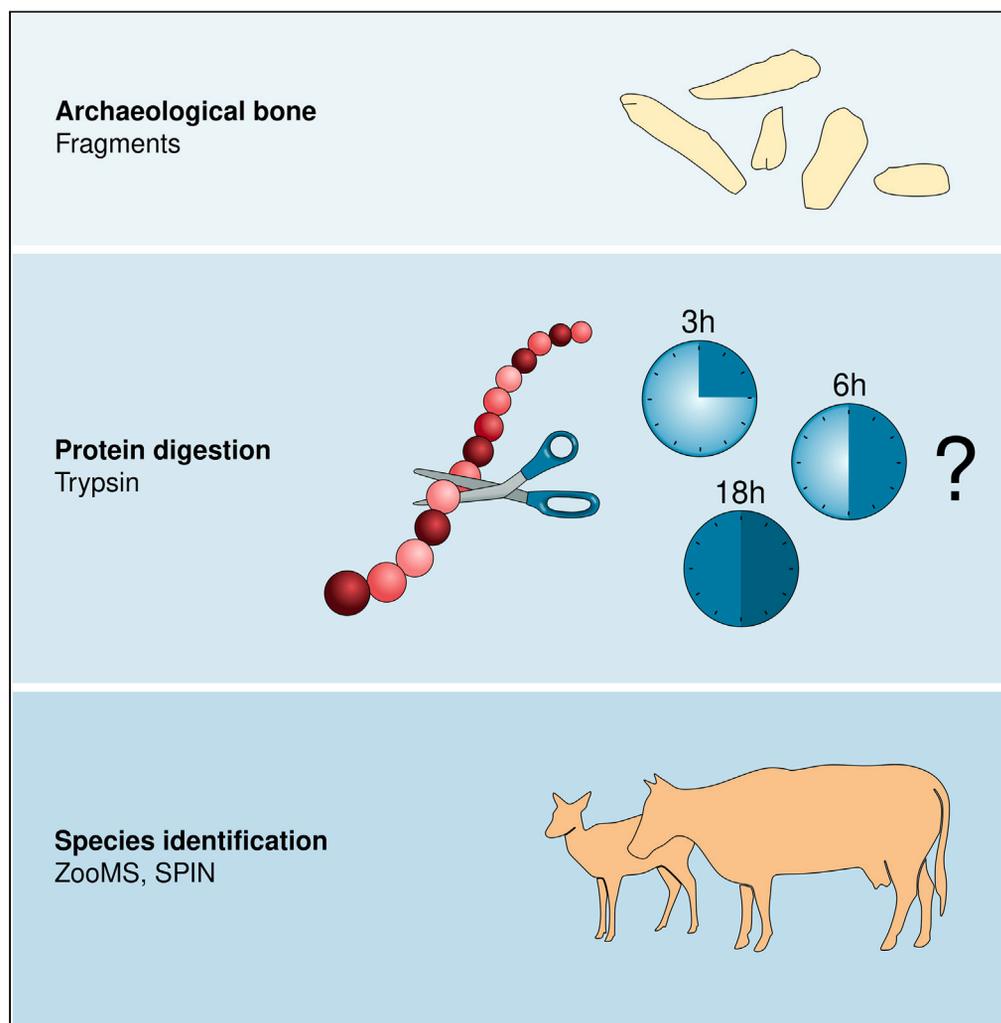**Article**

Increasing sustainability in palaeoproteomics by optimizing digestion times for large-scale archaeological bone analyses

Louise Le Meillour,
Virginie Sinet-Mathiot,
Ragnheiður Diljá
Ásmundsdóttir, ...,
Geoff M. Smith,
Zandra Fagernäs,
Frido Welker

frido.welker@sund.ku.dk

Highlights

Changes in laboratory protocols may have significant effects on environmental impact

Paleoproteomic approaches such as ZooMS or SPIN often process thousands of samples

Reduction of digestion duration has no measurable impact on species identification

Sample processing in plates instead of microtubes significantly reduces emissions

Article

Increasing sustainability in palaeoproteomics by optimizing digestion times for large-scale archaeological bone analyses

Louise Le Meillour,[1,8] Virginie Sinet-Mathiot,[2,3,8] Ragnheiður Diljá Ásmundsdóttir,[1] Jakob Hansen,[1,4] Dorothea Mylopotamitaki,[5] Gaudry Troché,[1] Huan Xia,[6] Jorsua Herrera Bethencourt,[1] Karen Ruebens,[5] Geoff M. Smith,[7] Zandra Fagernäs,[1] and Frido Welker[1,9,*]

SUMMARY

Palaeoproteomic analysis of skeletal proteomes is used to provide taxonomic identifications for an increasing number of archaeological specimens. The success rate depends on a range of taphonomic factors and differences in the extraction protocols employed. By analyzing 12 archaeological bone specimens from two archaeological sites, we demonstrate that reducing digestion duration from 18 to 3 hours has no measurable impact on the obtained taxonomic identifications. Peptide marker recovery, COL1 sequence coverage, or proteome complexity are also not significantly impacted. Although we observe minor differences in sequence coverage and glutamine deamidation, these are not consistent across our dataset. A 6-fold reduction in digestion time reduces electricity consumption, and therefore CO_2 emission intensities. We furthermore demonstrate that working in 96-well plates further reduces electricity consumption by 60%, in comparison to individual microtubes. Reducing digestion time therefore has no impact on the taxonomic identifications, while reducing the environmental impact of palaeoproteomic projects.

INTRODUCTION

The climate crisis that our planet is currently facing requires action from all parts of society in an attempt to mitigate its effects.[1] Simultaneously, the effect of climate change on the preservation of archaeological landscapes,[2–4] archaeological sites[5–7] and the recovery and preservation of individual heritage objects[8,9] is becoming increasingly evident.[10] Next to monitoring and predicting these effects, archaeological research therefore also has a responsibility in trying to reduce its impact on the planetary ecosystems. One aspect through which archaeological investigation can directly contribute is via lowering the emissions associated with laboratory work. Such activities are both energy-demanding and consume large quantities of single-use plastics. Even small adjustments to the way laboratories operate can significantly reduce the environmental impact; for example, increasing the temperature of ultra-low temperature freezers from −80°C to −70°C significantly reduces energy consumption[11] without affecting the stability of a range of biological materials.[12,13] Identifying and applying further adjustments can, cumulatively, significantly reduce the environmental impact of laboratory work.

The analysis of proteins preserved in archaeological materials is a fast-growing field, allowing for the taxonomic identification of skeletal remains[14–23] studying phylogenetic relationships between extinct and extant taxa[24–27] and identifying dietary elements in past societies,[28–30] amongst others. Approaches such as ZooMS (Zooarchaeology by Mass Spectrometry)[31] and SPIN (Species by Proteome INvestigation)[32] allow for high-throughput and medium-throughput processing, respectively, of a large number of bone specimens, even thousands of samples.[33] These large-scale proteomic investigations in particular are becoming a more routine part of (zoo)archaeological investigations. Due to their scale, they consume significant amounts of plastics, chemicals and electricity, especially when compared to the relatively smaller number of archaeological specimens subjected to full proteome analysis.

Most palaeoproteomics studies are bottom-up approaches with enzymatic digestion of solubilized proteins using trypsin, which requires heating to 37°C for optimal activity. Digestion is a critical step for the mass spectrometry analysis of proteins, but can be relatively time-consuming. Although there is no complete consensus across the field, many palaeoproteomic protocols include an overnight trypsin

[1]Globe Institute, University of Copenhagen, 1353 Copenhagen, Denmark
[2]University of Bordeaux, CNRS, Ministère de la Culture, PACEA, UMR 5199, 33615 Pessac, France
[3]University of Bordeaux, CNRS, Bordeaux INP, CBMN, UMR 5248 and Bordeaux Proteome Platform, 33076 Bordeaux, France
[4]Departament de Prehistòria, Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain
[5]Chaire de Paléoanthropologie, CIRB, Collège de France, Université PSL, CNRS, INSERM, 75005 Paris, France
[6]College of Earth and Environmental Sciences, Lanzhou University, Lanzhou 730000, China
[7]School of Anthropology and Conservation, University of Kent, CT2 7NRH Canterbury, UK
[8]These authors contributed equally
[9]Lead contact
*Correspondence: frido.welker@sund.ku.dk
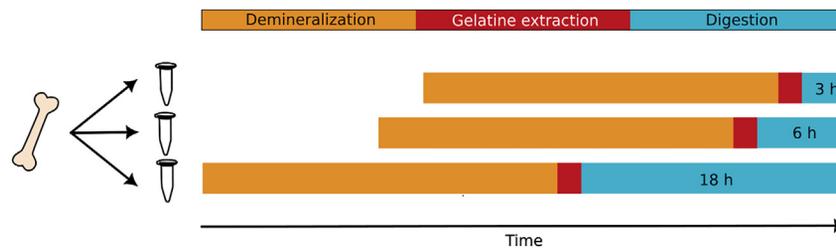https://doi.org/10.1016/j.isci.2024.109432

Figure 1. Schematic of the staggered laboratory workflow to test the impact of the different digestion times
Subsequent to digestion, the digests were acidified and split in two halves — one for ZooMS analysis and one for SPIN analysis. Note that the time arrow indicates real time, with all digestions terminated at the same time. As a result, subsequent steps such as peptide purification happened simultaneously for all three digestion conditions.

digestion of approximately 18 h,[24,32,33] sometimes under agitation (Table S1). However, studies of modern proteins have shown that an overnight digestion with trypsin, compared to shorter digestion times, leads to lower amino acid coverage.[34] Interestingly, there is no quantitative comparison on the relationship between trypsin digestion duration and the success of obtaining a taxonomic identification in high- and medium-throughput palaeoproteomic applications. The ability to reduce digestion duration would enable single-day sample preparation for ZooMS, enhancing throughput capacities even further. In addition, shorter digestion durations in high-throughput taxonomic identification studies in particular would also significantly reduce the environmental impact of palaeoproteomics research.

In this study, we explore the effects of reducing trypsin digestion time on palaeoproteomics analysis of herbivore skeletal elements from the Middle Pleistocene layers of Baishiya Karst Cave (China) and the Early Holocene layers of La Draga (Spain). Based on sedimentary and chronological differences between both sites, as well as the inclusion of specimens from a terrestrial and a phreatic/aquatic environment at La Draga, our specimens represent a range of preservation conditions. We compare digestion overnight (here defined as 18 h) to shorter digestion times of 6 h and 3 h, using both a MALDI-ToF MS and a LC-MS/MS approach, and interrogate the resulting data in terms of obtained taxonomic identifications, and proteome and protein degradation. If the taxonomic identifications are not negatively influenced by reducing digestion duration, this would favor reducing digestion durations in future palaeoproteomic studies in case taxonomic identification is the main purpose of the analysis, in order to minimize laboratory CO_2 emissions and their subsequent impact on global environmental conditions.

RESULTS

Twelve bone specimens were selected from two archaeological sites, La Draga (hereafter LD, Spain) and Baishiya Karst Cave (hereafter BKC, China; see STAR Methods section, Table S2). These sites were chosen because they cover a range of preservation environments and different chronologies. Each specimen selected contained cortical tissue, which was sampled for proteomic analysis. All of the sampled bones had prior taxonomic information, either obtained with ZooMS or through comparative anatomy. Including our proteomic observations, eleven specimens represent *Bos* sp./*Bison* sp. and one specimen is identified as a cervid (most likely as *Cervus elaphus*).

Preserved proteins from the twelve specimens were extracted using published protocols and sub-sampled for comparing three different digestion duration: 3, 6 and 18 h (for more detailed information, see STAR Methods section, Figure 1).

Reducing digestion duration has no effect on ZooMS and SPIN taxonomic identifications

Amongst the twelve samples analyzed in this paper, all allowed for taxonomic identification, regardless of the site (BKC or LD) and digestion duration (18, 6, and 3 h) through ZooMS. One sample from LD, namely specimen LD_02, showed peptide markers of Cervidae with ZooMS. This sample can be attributed to *Cervus elaphus* (red deer), while *Alces alces* (elk), *Dama* sp. (fallow deer), and *Saiga* sp. or *Megaloceros giganteus* (giant deer) can be excluded based on our knowledge of the fauna in this region and period. All other eleven samples were consistently identified as *Bos* sp./*Bison* sp. (Figure 2). Therefore, MALDI-ToF MS analysis allowed for the same level of ZooMS taxonomic identifications regardless of the digestion duration.

Comparatively, SPIN performed more heterogeneously across sites and digestion durations (Figure 2). SPIN assigned specimen LD_02 to Cervidae as well in all digestion durations, consistent with the observations made with ZooMS. A further 20 extracts were assigned to *Bos* sp./ *Bison* sp., consistent with the identifications provided by ZooMS. Although the SPIN taxonomic identification (the "Species" output field) is, in part, driven by the FineSpecies assignment, the BestMatch feature ranks *Bos mutus* (wild yak) beyond attribution to other Bovinae for six extracts deriving from four different bone specimens (Table S3). These four bone specimens all derive from BKC, making this assignment both likely as well as ecologically and climatically relevant.[35] Six extracts were exclusively assigned to the genus *Bos* sp., all from LD, including all three extracts from specimen LD_01. These taxonomic attributions are beyond what is possible with ZooMS. Interestingly, using the SPIN workflow five extracts were not assigned a taxonomic identification (twice for a 3 h digest, twice for a 6 h digest, and once for an 18 h digest). Moreover, two digests were assigned to a taxonomic group that, in addition to the genera *Bos* sp. and *Bison* sp. also included *Bubalus* sp. (Figure 2; Table S3). Given that the non-successful identifications are not specific to shorter or longer digestion durations, we conclude that digestion duration is not influencing the success of the SPIN taxonomic identifications.
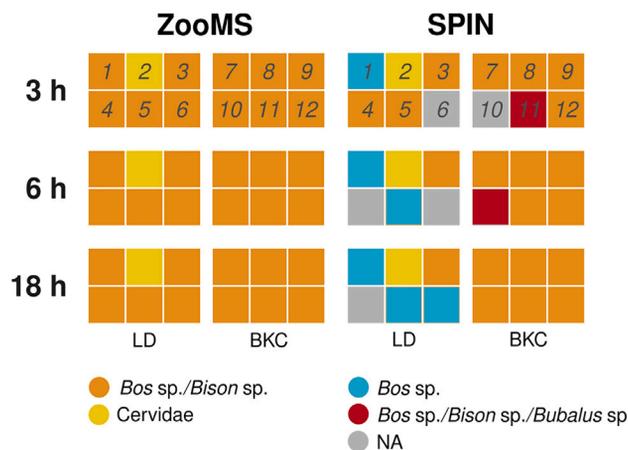
Figure 2. ZooMS and SPIN taxonomic identifications across the three digestion times
Specimen numbers shown in the first block are consistent across blocks and taxonomic identification methods (ZooMS analysis based on MALDI-ToF MS data acquisition and SPIN analysis based on LC-MS/MS data acquisition).

Effects on data quality for taxonomic identifications (ZooMS and SPIN) is negligible

Overall, both ZooMS and SPIN analyses presented well-preserved collagen type I sequences in at least one extract of every specimen analyzed. We observe a minimum of seven and a maximum of nine peptide markers observed for ZooMS (Figure 3). For SPIN, the minimum amino acid positions assigned to a positive identification was 596. At La Draga, some extracts recovered close to 2,000 amino acid positions, including up to 400 amino acid positions deriving from non-collagenous proteins, among the 20 proteins considered within the SPIN analysis (Figure 3). We observe some differences between the archaeological contexts considered (BKC, LD Sector A and LD Sector B) in terms of the SPIN amino acid counts and the ZooMS peptide marker counts recovered. In reference to the ZooMS nomenclature,[36] we note the absence of peptide marker α2 454 within all three conditions (absent in 83% of the spectra for the 3 h trypsin digestion, 58% for the 6 h trypsin digestion and 66% for the 18 h trypsin digestion; Table S4). We also note the absence of peptide marker α2 502 in LD_01 (6 h) and peptide marker α2 978 in samples BKC_12 (3 h) and BKC_12 (6 h). The absence of these peptide markers does not influence the taxonomic assignment of the samples, and the reduction of the trypsin digestion duration is not associated with a statistically significant difference in the number of identified peptide markers (Kruskal Wallis test, n.s. for both sites).

Furthermore, the mean number of monoisotopic peaks, retrieved from the MALDI-ToF MS data were calculated across samples for each digestion duration (Figure S1). The number of peaks is not significantly affected by digestion duration, but rather by the archaeological context (ANOVA type II: F = 16.17, p = 0.01). Additionally, the intensities (Figure S2) across all the monoisotopic peaks detected in our MALDI-ToF MS data were significantly different across digestion durations (ANOVA type II: F = 4.54, p = 0.018). Intensities are generally higher at the digestion duration of 18 h compared to 6 h and 3 h within all sites. When considering the three main peptides (α1 508, α2 484 and α2 793) from each extract obtained through MALDI-ToF MS, signal-to-noise ratios appear to vary between digestion durations, but are only statistically significant for peptide α2 484 (ANOVA type II: F = 4.19, p = 0.024; Figure S3). A similar pattern is noticeable for the absolute peak intensity, with the digestion duration being a factor variable but only significantly in the case of α2 793 (ANOVA type II: F = 3.45, p = 0.044; Figure S4). Therefore, it appears that although there are differences in absolute intensities and signal-to-noise ratios, these are heterogeneous and have not influenced taxonomic specificity of the obtained ZooMS identification in a negative manner.

In addition to the quality of our MALDI-ToF MS spectra, we calculated a range of measures for the SPIN analysis of each extract (Table S5). We generally observe comparable numbers of acquired and identified MS2 scans across digestion durations (ANOVA type II: n.s. and ANOVA type III: n.s., respectively; Figure S5), although rates of identified MS2 spectra might be slightly higher for 3 h digestion durations, while the number of MS2 scans acquired might be slightly higher, on average, for 18 h durations. Importantly, the number of proteins identified in total (ANOVA type III; n.s.), as well as separately for collagenous (ANOVA type III; n.s.) and non-collagenous proteins (hereafter, NCPs, ANOVA type II; n.s.), remains consistent across digestion durations (Figure S6). However, in the context of the SPIN protein sequence database we find that digestion duration does not significantly affect the number of recovered sites (ANOVA type III: n.s.), with the highest number being recovered through 18 h digestion duration for LD and 6 h digestion duration for BKC (Figure S6). NCPs were generated for several LD specimens for all digestion times, but only for one 3 h digested BKC specimen. The maximum number of collagen proteins identified in our SPIN LC-MS/MS data was 12 (out of 12 possible), regardless of the digestion duration considered (ANOVA type III: n.s.; Figure S6B). Even though samples of 3 h digestion generated the highest number of collagen proteins in the LD samples, we obtained the lowest number for BKC using this digestion duration.

Without taking the presence and placement of potential PTMs into account, the number of unique amino acid sequences was comparable for specimens recovered within archaeological contexts, regardless of the digestion duration (ANOVA type III: n.s., Figure S7B). We recovered the highest number of unique peptides in the 18 h digestion samples of LD and in the 6 h digested samples of BKC (Figure S7B). The same
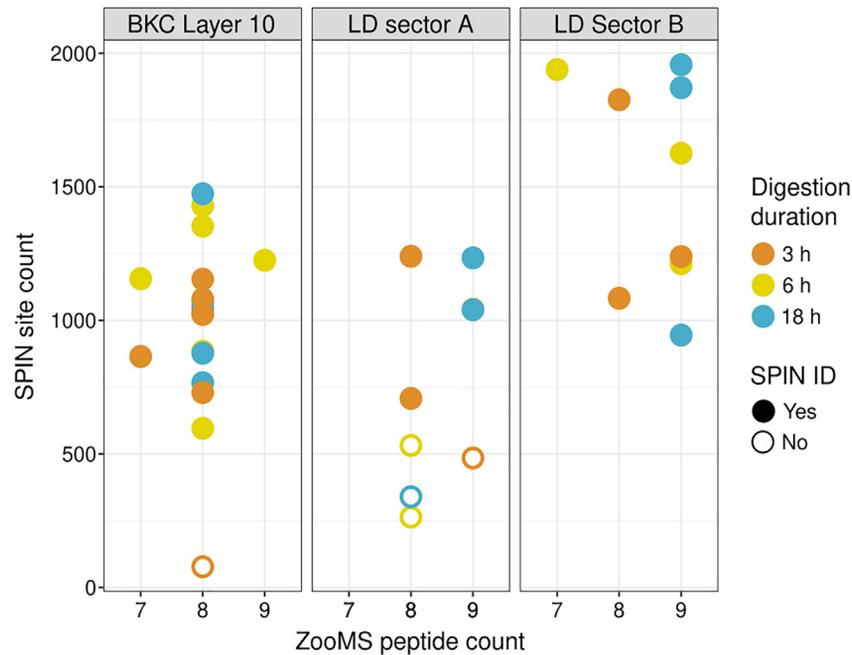
Figure 3. Peptide markers obtained for ZooMS compared to site counts obtained for SPIN
Points without fill represent extracts where SPIN did not provide a taxonomic identification (NA).

pattern is observed for the amino acid sequence coverage of NCPs. Here, only one BKC specimen processed (extract 08_3h) generated a number of NCP site counts for a single protein, while for LD there were no significant differences in NCP site counts between digestion durations (ANOVA type II: n.s.; Figures S6C and S7D). For collagens, the 18 h digestion duration generated the highest number of site counts; however, there were no significant differences among the different digestion times (ANOVA type III: n.s.; Figure S7C). Furthermore, we observe that roughly 75% of the identified peptides have no missed cleavages internally, regardless of digestion duration (Figure S8). Based on these observations of obtained site counts and their distribution across the 20 proteins considered in SPIN analysis, there appears to be no consistent relationship between any of these measures and digestion duration.

The preceding has shown that some statistical comparisons are significantly different between digestion durations within an archaeological context, but that the absolute differences are never large. Moreover, significant differences in one parameter at one archaeological context are never replicated with the same effect in another archaeological context. Therefore, digestion duration appears to have little effect on the ability of ZooMS and SPIN to assign taxonomic identities to extracts of Holocene and Pleistocene skeletal proteomes.

Proteome composition is largely influenced by archaeological contexts

The data search conducted against the whole reference proteome of *Bos taurus* allowed for the identification of 42 different protein groups (after filtering for potential contaminants). These proteins overlap in part with those identified through the SPIN-specific database search, and include a range of collagens as well as NCPs. All these proteins have been identified in skeletal proteomes before. Collagenous and non-collagenous proteins (NCPs) were identified from both sites for all three digestion durations. The number of proteins, both collagenous and NCPs, increases or remains similar between the different digestion durations (Figure 4A). One exception lies in the number of NCPs present in the BKC samples, where the average number of these proteins varies from 1.5 for 3 h, to 2.5 for 6 h, and to 1.4 for 18 h. No significant effect on the number of collagenous or non-collagenous proteins identified was observed (ANOVA type III: n.s.) in relation to digestion duration or archaeological context. The near-absence of NCPs in the Baishiya Karst Cave specimens is consistent with the previous observation that surviving proteomes at the site are almost completely composed of collagenous proteins.[37]

When taking into account the entire *Bos* proteome, no significant difference was observed relating to the effect of digestion duration on the number of peptides identified (ANOVA type II: n.s.; Figure 4B). This is consistent with our observations for the SPIN data analysis reported above. In contrast, the estimated number of amino acids covered, i.e., the number of uniquely identified amino acid positions for a protein group, seems to increase with longer digestion time (Figures 4C, S9, and S10). In terms of absolute protein sequence coverage, there might therefore be an advantage to digest for longer periods of time. Our SPIN data analysis indicates this will not necessarily improve taxonomic assignments, however, as demonstrated above, there is no significant relationship between the NCP site count or collagens site count recovered through SPIN data analysis (Figure S7). It would therefore seem that the additional protein sequence coverage obtained through longer digestion durations does not significantly impact the taxonomic assignment of such proteomes.
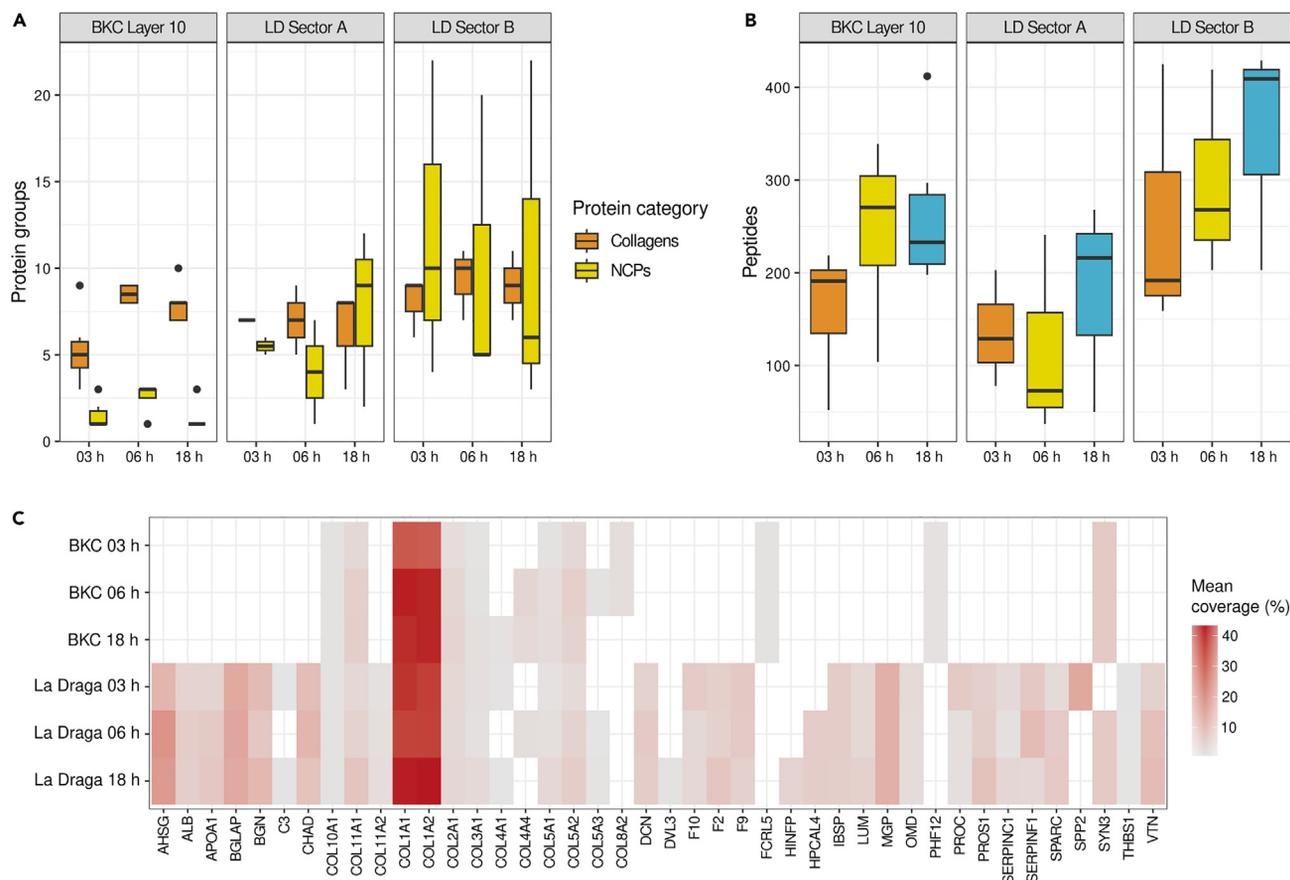
Figure 4. Characteristics of the full proteomes

(A) Number of identified proteins (separated into collagens and NCPs), (B) Number of recovered peptides, (C) Mean amino acid coverage per protein across samples. Note that sequence coverage is expressed as a mean percentage in comparison to the full length of the proteomic entry in UniProt, not the length of the mature and secreted protein.

For the LD samples, the estimated number of amino acids covered for individual proteins ranges from 8 to 803 for 3 h, 8 to 728 for 6 h, and 8 to 810 for 18 h digestion (Figure S9). For the BKC samples, the same estimates range from 8 to 566 for 3 h, 8 to 699 for 6 h, and 8 to 768 for 18 h digestion (Figure S10). No significant relationship was detected between the estimated number of amino acids covered and the digestion duration (ANOVA type III: n.s.). However, a significant relationship between the archaeological context of the specimens and the estimated number of amino acids covered was present (ANOVA type III: F = 8.98, p < 0.001).

The archaeological contexts from which our 12 specimens derive vary in terms of chronological age (BKC versus LD) and sedimentary environments, with La Draga specimens deriving either from Sector A, which represents a terrestrial environment, or Sector B, which represents a more phreatic or aquatic environment. Unsurprisingly, we therefore observe more protein groups in the three digestion durations for the LD samples compared to the BKC samples (Figure 4A). We further determine no significant differences in proteome composition when comparing the numbers of collagenous and non-collagenous proteins (ANOVA type III: n.s.) to digestion duration or archaeological context of specimens for our full *Bos* proteome analysis.

Contamination and proteome degradation

To assess any potential contamination by non-endogenous peptides, we extracted laboratory blanks alongside the samples. In the MALDI-ToF MS spectra, these remained empty of collagenous peptides, excluding the possibility of considerable laboratory contamination. In the SPIN analysis, we observe some collagen peptides (site counts ranging between 12 and 228 for collagens) in the extraction blanks, which is the primary reason why some archaeological samples were not assigned a taxonomic identity. In addition, we observe three peptides resulting from trypsin autolysis within the MALDI-ToF MS blank spectra (Table S6) and note higher absolute intensities of their peptide peaks and higher signal-to-noise ratios in the case of a 3 h enzymatic digestion of the blanks when compared to 6 h and 18 h digests. This is consistent with our observation that 3 h digests are nearly entirely composed of trypsin in the blanks, while longer digestion durations have slightly lower relative protease ratios (Figure S11). Based on these observations, we determine that both our MALDI-ToF MS and LC-MS/MS data results and assigned taxonomic identities are not influenced by protein contamination.
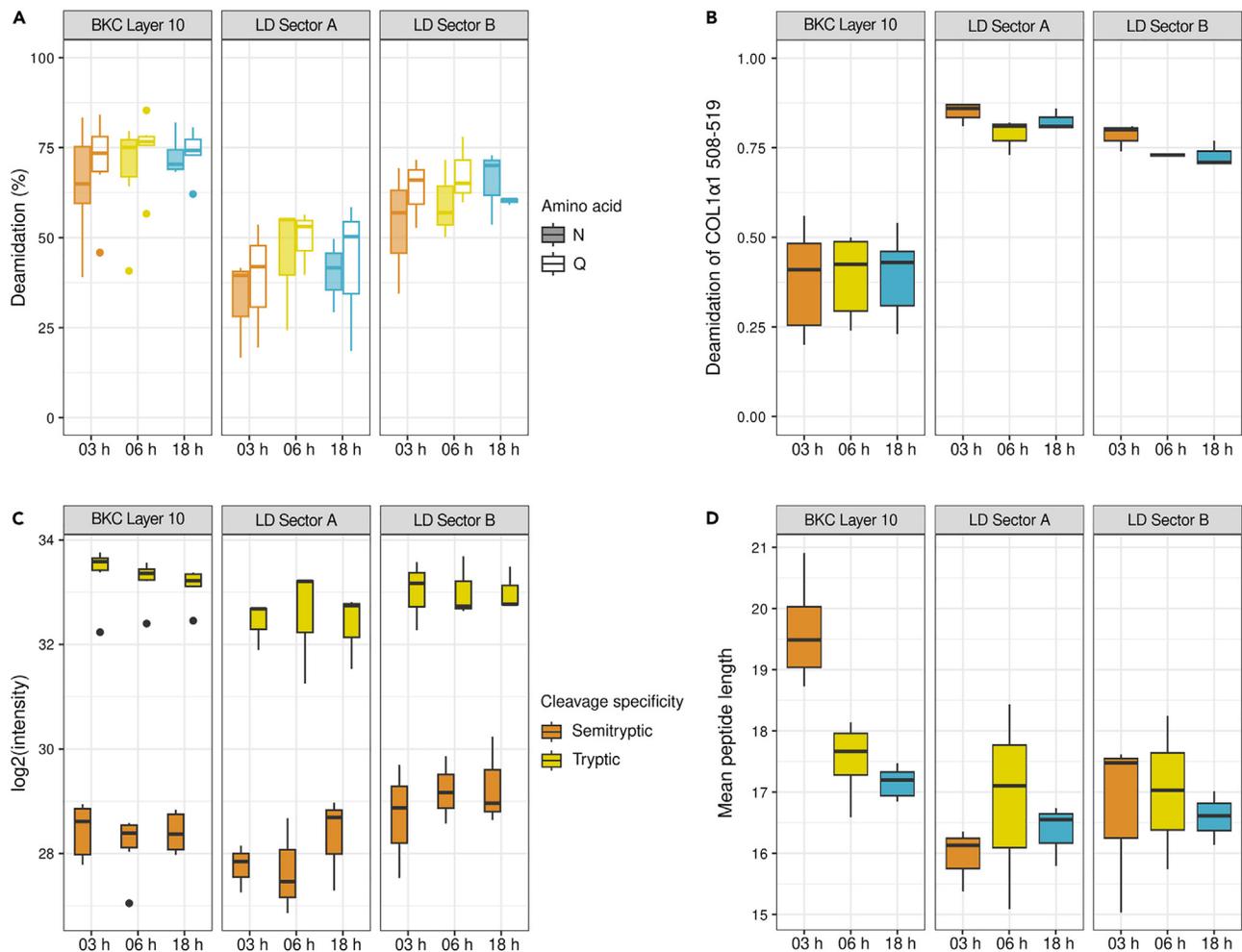
Figure 5. Proteomic data quality

(A) Deamidation, where 0% indicates no deamidation and 100% full deamidation of the proteome, with filled bars representing deamidation of N and non-filled bars representing deamidation of Q, (B) Deamidation of the COL1α1 508–519 peptide, with 0 indicating full deamidation and 1 indicating no deamidation of the relevant glutamine (Q), (C) Cleavage specificity, (D) Peptide length, weighted by intensity. Panels (A), (C), and (D) are based on the full *Bos taurus* proteome data, whereas (B) is based on the ZooMS data.

In addition, we determine elevated ratios of protein degradation as the result of diagenetic processes. Deamidation rates of asparagine (N) and glutamine (Q) are complex and can depend on various factors including pH, temperature, the characteristics and structure of the protein, but also the choice of methods for sample preparation.[38] Generally, under physiological conditions, asparagine deamidation tends to occur at a faster rate compared to glutamine deamidation.[39] Observations made on the rate of glutamine and asparagine deamidation in previous palaeoproteomic studies has shown that, in general, glutamine has a lower extent of deamidation compared to asparagine within the same specimen.[40,41]

Both MALDI-ToF MS and LC-MS/MS datasets reveal significant differences in deamidation between glutamine and asparagine deamidation, estimated by considering the percentage of glutamine and asparagine deamidation within the (full) proteome in the case of LC-MS/MS analysis and through the glutamine deamidation of the peptide marker COL1α1 508–519 in the context of MALDI-ToF MS analysis. Overall, we observe a more extreme deamidation for BKC compared to LD (Figures 5A and 5B; asparagine; ANOVA type III: F = 16.51, p < 0.001; Figure 5A). Our full proteome analysis reveals that asparagine deamidation is only affected by the archaeological context and not by digestion time, while glutamine deamidation is influenced by both archaeological context and digestion time (glutamine; ANOVA type III: F = 3.27, p = 0.036; Figure 5A). Specifically for glutamine, our dataset suggests that shorter digestion times result in lower extents of glutamine deamidation (Figure 5A). In the deamidation ratios obtained for one peptide observed in our MALDI-ToF MS dataset, COL1α1 508–519, we likewise find significant differences in deamidation between our three preservation contexts (ANOVA type II: F = 144.83, p < 0.001). In this case, however, there appears to be no significant influence of digestion duration on the glutamine deamidation value obtained for the single amino acid that can deamidate in this peptide (ANOVA type II: n.s.).

In addition to glutamine and asparagine deamidation, we looked at the prevalence of semitryptic versus tryptic peptides (Figure 5C) and the peptides length (Figure 5D). For cleavage specificity, we observe that the total intensity of both semitryptic and tryptic peptides is not affected by digestion duration, but that the observed values for both are dependent on the archaeological context (ANOVA type III: F = 8.43, p = 0.001 and F = 6.15, p = 0.006, respectively; Figure 5C). In contrast, mean peptide length, weighted by the intensity of the peptide, differs significantly between archaeological sites (ANOVA type II: F = 30.32, p < 0.001) and digestion durations (ANOVA type II: F = 5.97, p = 0.007). A significant interaction between the two factors is present (ANOVA type II: F = 10.13, p < 0.001). This is particularly noticeable in the BKC extracts, where a shorter digestion duration results in longer peptide lengths (Figure 5D).

The archaeological context from which our 12 bone specimens derive is therefore the primary driver of differences in proteome composition and its modification, regardless of the duration of trypsin digestion. The exception to this is glutamine deamidation, where some glutamine sites might be prone to deamidate in the case of longer digestion durations, and peptide lengths, weighted by intensity, where shorter digestion durations might result in a larger proportion of longer peptides. As the latter is observed for the BKC proteomes that are largely composed of collagen, this observation might be particularly noteworthy for phylogenetic studies of skeletal proteomes dominated by this protein group. However, as both our SPIN data analysis and the estimated amino acid coverage for the *Bos* proteome search indicated, there is no difference in taxonomic identification of these extracts.

The environmental impact of palaeoproteomics can be significantly reduced

Electricity consumption and its emission intensity can be reduced by a factor of 6 when decreasing digestion time from 18 h to 3 h (Figure 6; Table S7). In doing so, it is possible to decrease the gCO$_2$eq per extract from 3.79 ± 2.41 SD to 0.66 ± 0.40 SD for tubes, and from 1.69 ± 2.38 SD to 0.28 ± 0.40 SD for plates based on the carbon intensities from countries included in this study. In changing from microtubes to 96-well plates, it is further possible to decrease gCO$_2$eq per extract by 62.85% at 18 h and by 70.15% at 3 h digestion duration. Additionally, it is worth mentioning the large difference in gCO$_2$eq emissions between countries, further impacting the relative gCO$_2$eq emission. The gCO$_2$eq emitted per kWh consumed differs by geographic region, here country averages are used, depending on the energy sources used within that region to generate electricity. As a result, geographic regions where energy is largely derived from low-carbon resources such as wind, solar, or nuclear energy, will be associated with lower gCO$_2$eq per kWh of electricity consumed compared to geographic regions where high-carbon resources, such as coal and gas, contribute most to the generated electricity.[42] Therefore, by changing from microtubes to 96-well plates, decreasing the digestion time from 18 h to 3 h, and depending on the country of analysis, gCO$_2$eq emissions can be decreased significantly (linear model, F = 8.31 + 18, p < 0.001).
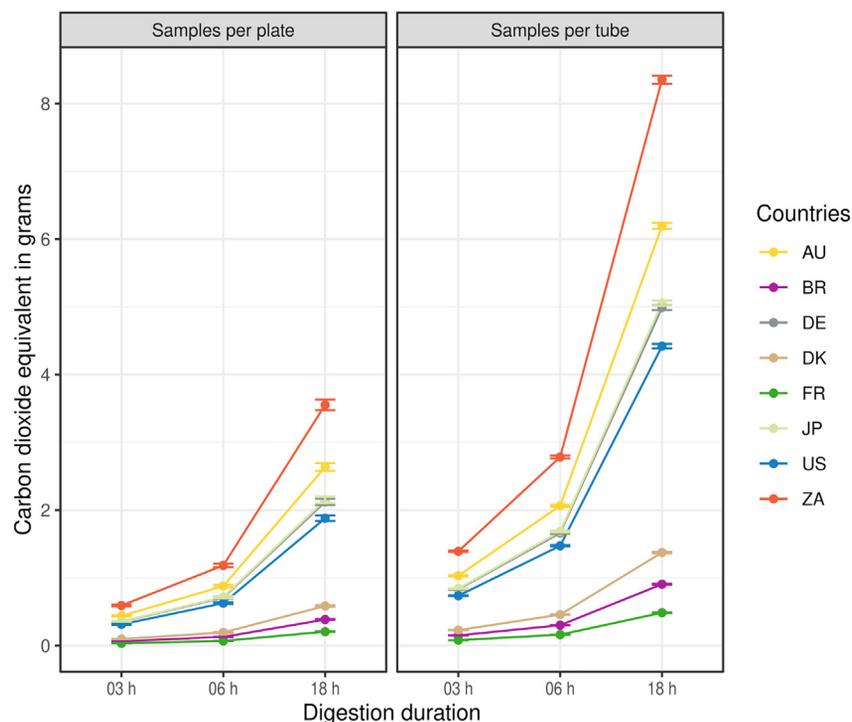

Figure 6. Carbon dioxide emission in grams per digestion, for the three different digestion durations performed using either individual microtubes or 96-well plates
Error bars are ±2 SD. Australia: AU, Brazil: BR, Germany: DE, Denmark: DK, France: FR, Japan: JP, USA: US and South Africa: ZA.

DISCUSSION

Academia can address the current questions around sustainable practices and the global climate crisis through offering solutions applicable to society and/or by changing its own practices.[43] Although the impact of individual researchers or fields of research might appear minimal on a global scale, even small adjustments to the way laboratories operate can significantly reduce the environmental impact of their research locally.[44] For example, an approach that has recently gained attention is increasing the temperature of many ultra-low temperature freezers from $-80°C$ to $-70°C$. This temperature change significantly reduces energy consumption,[11] and thereby both costs as well as emissions, without affecting the stability of a range of biological materials.[12,13] Calls of action toward self-reflection and incentivizing sustainable academic practices, including laboratory practices, have existed for a long time, but these have generally been slow to be implemented.[45–48]

In this context, we observe that over the past two decades an increasing number of archaeological and paleontological skeletal specimens are studied through protein mass spectrometry methods. The largest proportion of these specimens are analyzed through ZooMS or SPIN, with the sole purpose of providing a taxonomic identification.[32,49,50] Cumulatively, therefore, it can be expected that it is the extraction of these thousands, if not tens of thousands, of bone proteomes that utilize the most electricity during the protein extraction process.

As with many extraction protocols, interventions could be made at several protocol steps in order to minimize electricity consumption, or reduce consumption of chemical and plastic consumables. Trypsin digestion duration at $37°C$ is one such aspect, being the longest step in many published protocol descriptions (Table S1). Although previous proteomic studies have indicated that digestion duration can have an impact on proteome composition, and protein sequence coverage, comparative studies on this aspect have so far been absent in palaeoproteomics.[34,51–53]

Implications of reducing digestion duration on the level of taxonomic identification

The results presented above indicate that reducing digestion duration in palaeoproteomics protocols has no significant effect on our ability to provide a taxonomic identification of an archaeological specimen, at least in the context of ZooMS or SPIN analysis. The reduction of the digestion duration provides analyzable MALDI-ToF MS spectra and allows the retrieval of peptide markers across the m/z range taken into account for ZooMS, allowing the taxonomic identification of the samples at the expected resolution. Similarly, for the same extracts, useful LC-MS/MS datasets can be obtained. There is a correct correlation between taxonomic identities previously obtained and the ones presented in this study, except for one specimen, which was re-attributed to Cervidae based on ZooMS and SPIN taxonomic attributions. Taking into account the bio-chronological context of the specimen and relevant species present in Holocene Iberia allowed specimen LD_02 to be identified as a red deer (*Cervus elaphus*).

Whenever taxonomic identities were assigned in both ZooMS and SPIN analysis, the two approaches provide compatible identifications. As in a previous study, we observe that SPIN in some cases provides taxonomic identifications beyond the limits present in ZooMS.[32] In the current case, this is visible as the identification of some specimens to the genus *Bos* specifically, excluding the possibility of an attribution to *Bison*. Surprisingly, though, we also observe several cases where the SPIN identifications are less precise than the ones provided by ZooMS, and extracts that do not result in an identification at all (Figure 2). As such less-specific taxonomic identifications are not more prevalent in one digestion duration compared to another, we conclude that digestion duration has no impact on the achieved taxonomic identifications for ZooMS or SPIN.

Interestingly, we were unable to identify a significant impact of digestion duration on many of the parameters we tested. This appears true for proteome composition and peptide numbers identified which are relevant aspects when attempting a taxonomic classification. We determined that deamidation of some glutamine amino acid sites might be less advanced when digesting for shorter durations, especially when observed through LC-MS/MS analysis, but were unable to observe this in our ZooMS dataset. The quantification of diagenetic post-translational modifications, such as deamidation, is becoming an important aspect of palaeoproteomic studies, especially in the context of determining and controlling for the presence of exogenous contaminants.[25,54,55] We therefore argue that researchers should standardize, and report, the trypsin type, concentration, and digestion duration they use in their study, especially when making comparative claims in regards to other studies.

Effects of digestion duration on lab-based experiment energy consumption

The reduction of the digestion duration allows for a considerable reduction of the electricity consumption and therefore of the emission intensity. Furthermore, per-extract emissions can be reduced even further when working in 96-well plates in comparison to microtubes. Considering that reducing digestion duration 6-fold has no negative impact on the achieved taxonomic identifications, this further reduction in emissions would be associated with increased sample throughput and shorter laboratory processing times overall. Further work should quantify whether this positive benefit of 96-well plates, in comparison to microtubes, is in part offset by increased plastics production costs. Both the implementation of a reduced digestion duration as well as a switch to working in 96-well plates are easy to set up in most biomolecular laboratories used for palaeoproteomic studies, and our observations could therefore have widespread positive benefits for the emissions produced by the palaeoproteomics research field.

Future prospects for increasing sustainability in the field of ancient protein analysis

We focused our study on a single step, protein digestion, in the common ZooMS and SPIN workflows where the taxonomic identification of the resulting skeletal proteome is the primary research objective. This step does not exist in isolation. Future research on

increasing the sustainability of palaeoproteomic research can, and maybe should, consider aspects related to the logistics around specimen and peptide extract transportation, emissions associated with the use of disposable plastics, the location (Figure 6), and the type of mass spectrometry instrumentation used. It can be expected that mass spectrometry instrumentation, for example, produces significantly larger emission intensities compared to protein digestion. However, some of these aspects such as minimizing the emission intensities of protein mass spectrometry instrumentation might be outside of the control of many practitioners in the field, while sample preparation, including protein extraction, digestion, and purification, generally is within the control of people performing these types of protein extractions. There should be further opportunities to decrease the emission intensity associated with large-scale palaeoproteomics research within our laboratories, for example around the use of plastics, chemicals, and other disposable consumables, or the simplification of extraction steps[56], in line with current developments of sustainable, efficient, and economically viable practices in molecular laboratories[57–59] and the realization that climate change negatively impacts archaeological landscapes, sites, and cultural heritage objects.[10]

The exploration of sustainable practices should be guided by the understanding that the data quality in relation to the research objectives should not be negatively impacted. In our study, we observe that taxonomic identifications do not become less precise at shorter digestion durations, and that many quality parameters show no relationship with digestion duration either. Simultaneously, we observe that overall protein sequence coverage across the entire proteomes decreases slightly at shorter digestion durations. Therefore, to further optimize the digestion duration of taxonomic studies in palaeoproteomics, we call for researchers to consistently report the duration, temperature, catalog number, and supplier of any protease used (Table S1). In addition, we note that alternative trypsin proteases exist that are claimed to enhance digestion efficiency in association with shorter digestion durations. Although these products have not been widely used in the palaeoproteomics field, and would require benchmarking against commonly used proteases before general adoption, they may provide an adequate solution to retain absolute sequence coverage at short digestion durations.[60,61]

We set out to determine whether trypsin digestion duration can be significantly reduced in high-throughput palaeoproteomic studies, when the ultimate aim is the taxonomic identification of a skeletal proteome. We find that reducing digestion duration from 18 to 3 h does not negatively affect the success rate of obtaining a taxonomic identification, a conclusion that appears valid for both ZooMS (MALDI-ToF MS) and SPIN (LC-MS/MS) datasets generated from the same extracts. We replicate previous findings that SPIN enables more precise identifications compared to ZooMS in some cases, but also find that in other cases SPIN analysis was not able to assign any taxonomic identity, a result absent from our ZooMS dataset.

We determine that electricity consumption and emission intensity can be significantly reduced when switching to 3 h digestion durations, resulting in a 6-fold reduction in emission intensity, while switching to working in 96-well plates instead of microtubes reduces emission intensity by a further 60%. A shorter digestion duration also enables ZooMS and SPIN extraction workflows to be performed in one or two days, instead of the current two or three days, depending on the protocol used. In addition to the benefits that this approach has for global environmental wellbeing, our findings therefore also have positive implications for laboratory efficiency. If the digestion time can be adapted so that the extraction protocol fits within a single working day, the wellbeing of laboratory workers will also be improved through an easier scheduling of laboratory sessions and time management within a working week. We thus argue that digestion duration should be significantly reduced in palaeoproteomics analyses when taxonomic identification is the main purpose of the analysis.

Finally, our analysis demonstrates that sustainable practices can have a positive effect on the emission intensities associated with ancient molecular research. We envision that further steps can be taken to further improve sustainable practices along the process from experimental design to publication, such as those associated with the use of single-use plastics, or those associated with the use of analytical instrumentation, without negatively impacting data quality. Given the growth of ancient molecular analysis over the past decades, we encourage our colleagues to find ways of reducing emission intensities associated with these activities, particularly considering the increasing evidence that climate change is negatively impacting archaeological heritage.

Limitations of the study

The experimental design is representative of the variability of palaeoproteomics projects that are conducted worldwide: two different archaeological sites, with different chronologies and various sedimentary context; three digestion durations; two different mass spectrometers and approaches to data analysis; and includes countries where this type of research is conducted into the electricity consumption measurement and comparison.

However, we acknowledge that many more parameters may affect the results. For example, only two archaeological sites are represented, and the archaeological (or paleontological) context from which the analyzed specimens originate might greatly influence protein retrieval, their diversity and diagenetic stage. Another parameter is the digestion duration: only three options were compared here, but further increases or decreases in duration may affect the results. We did not further monitor electricity consumption of the instruments used in our analysis (mass spectrometers), but acknowledge that this should count, to a large extent, in electricity consumption. The country in which we performed the analyses might also represent an important bias, both in terms of electricity production, consumption and subsequent CO_2 emissions. Overall, this study is a first attempt at monitoring and evaluating electricity consumption and CO_2 emissions of a specific aspect of palaeoproteomics analyses. We therefore show that the digestion step of most ancient protein analyses protocols could be reduced without significantly affecting the results of such studies.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- METHOD DETAILS
  - Sample selection
  - Bone proteome extraction
  - Mass spectrometry analyses
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - MaxQuant search
  - Data and statistical analysis
  - Electricity consumption and emission intensity

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.109432.

AUTHOR CONTRIBUTIONS

L.L.M., V.S.-M., Z.F., and F.W. designed the study. J.H., R.D.A., G.T., and F.W. conducted experiments. V.S.-M., R.D.A., J.H., D.M., X.H., J.H.B., and Z.F. conducted data analyses. G.M.S., K.R., Z.F., and F.W. contributed reagents, software or granted access to instruments. L.L.M., V.S.-M., Z.F., and F.W. wrote a first draft of the paper with input of all the authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Lee, H., and Romero, J. (2023). IPCC Synthesis Report Climate Change. https://www.ipcc.ch/report/ar6/syr/downloads/report/IPCC_AR6_SYR_LongerReport.pdf.

2. Ashton, N., Lewis, S.G., De Groote, I., Duffy, S.M., Bates, M., Bates, R., Hoare, P., Lewis, M., Parfitt, S.A., Peglar, S., et al. (2014). Hominin footprints from early Pleistocene deposits at Happisburgh, UK. PLoS One 9, e88329.

3. Erlandson, J.M. (2012). As the world warms: rising seas, coastal archaeology, and the erosion of maritime history. J. Coast Conserv. 16, 137–142.

4. Dawson, T., Hambly, J., Kelley, A., Lees, W., and Miller, S. (2020). Coastal heritage, global climate change, public engagement, and citizen science. Proc. Natl. Acad. Sci. USA 117, 8280–8286.

5. Tamantini, S., Sidoti, G., Antonelli, F., Galotta, G., Moscatelli, M.C., Kržišnik, D.,

Vinciguerra, V., Marabottini, R., Macro, N., and Romagnoli, M. (2023). The WOODPDLAKE Project. Lakes, Wood and Sediment: Natural and Cultural Heritage Affected by Climate Changes (EGU General Assembly 2023). EGU23-15116. https://doi.org/10.5194/egusphere-egu23-15116.

6. High, K., Milner, N., Panter, I., Demarchi, B., and Penkman, K.E.H. (2016). Lessons from Star Carr on the vulnerability of organic archaeological remains to environmental change. Proc. Natl. Acad. Sci. USA 113, 12957–12962.

7. Boethius, A., Kjällquist, M., Magnell, O., and Apel, J. (2020). Human encroachment, climate change and the loss of our archaeological organic cultural heritage: Accelerated bone deterioration at Ageröd, a revisited Scandinavian Mesolithic key-site in despair. PLoS One 15, e0236105.

8. Taylor, W.T.T., Dixon, E.J., Hafner, A., and Hinz, M. (2021). New directions in a warming world. J. Glacial Archaeol. 5, 1–3.

9. Pilø, L., Finstad, E., Ramsey, C.B., Martinsen, J.R.P., Nesje, A., Solli, B., Wangen, V., Callanan, M., and Barrett, J.H. (2018). The chronology of reindeer hunting on Norway's highest ice patches. R. Soc. Open Sci. 5, 171738.

10. Miller, S.E., and Wright, J.P. (2023). Introduction: Archaeology of the Anthropocene: Historical Archaeology's Response to the Climate Crisis. Hist. Archaeol. 57, 448–472.

11. Leak, L.B., Tamborski, J., Commissaris, A., and Brophy, J.A.N. (2023). Forging a path toward a more sustainable laboratory. Trends Biochem. Sci. 48, 5–8.

12. Corsini, J., Maxwell, F., and Maxwell, I.H. (2002). Storage of Various Cell Lines at -70°C or -80°C in Multi-Well Plates While Attached to the Substratum. Biotechniques 33, 42–46.

13. Beekhof, P.K., Gorshunska, M., and Jansen, E.H.J.M. (2012). Long term stability of paraoxonase-1 and high-density lipoprotein in human serum. Lipids Health Dis. 11, 53.

14. Sinet-Mathiot, V., Smith, G.M., Romandini, M., Wilcke, A., Peresani, M., Hublin, J.J., and Welker, F. (2019). Combining ZooMS and zooarchaeology to study Late Pleistocene hominin behaviour at Fumane (Italy). Sci. Rep. 9, 12350.

15. Le Meillour, L., Zazzo, A., Lesur, J., Cersoy, S., Marie, A., Lebon, M., Pleurdeau, D., and Zirah, S. (2018). Identification of degraded bone and tooth splinters from arid environments using palaeoproteomics. Palaeogeogr. Palaeoclimatol. Palaeoecol. 511, 472–482.

16. Charlton, S., Alexander, M., Collins, M., Milner, N., Mellars, P., O'Connell, T.C., Stevens, R.E., and Craig, O.E. (2016). Finding Britain's last hunter-gatherers: A new biomolecular approach to 'unidentifiable' bone fragments utilising bone collagen. J. Archaeol. Sci. 73, 55–61.

17. Welker, F., Soressi, M., Rendu, W., Hublin, J.-J., and Collins, M. (2015). Using ZooMS to identify fragmentary bone from the Late Middle/Early Upper Palaeolithic sequence of Les Cottés, France. J. Archaeol. Sci. 54, 279–286.

18. Buckley, M. (2018). Zooarchaeology by Mass Spectrometry (ZooMS) Collagen Fingerprinting for the Species Identification of Archaeological Bone Fragments. In Zooarchaeology in Practice (Springer), pp. 227–247.

19. Ebel, E., LeMoine, G.M., Darwent, C.M., Darwent, J., and Kirby, D.P. (2023). Using bone technology and ZooMS to understand indigenous use of marine mammals at Iita, Northwest Greenland. J. I. Coast Archaeol. 1–22.

20. Rey-Iglesia, A., de Jager, D., Presslee, S., Qvistgaard, S.S., Sindbæk, S.M., and Lorenzen, E.D. (2023). Antlers far and wide: Biomolecular identification of Scandinavian hair combs from Ribe, Denmark, 720–900 CE. J. Archaeol. Sci. 153, 105773.

21. McGrath, K., Rowsell, K., Gates St-Pierre, C., Tedder, A., Foody, G., Roberts, C., Speller, C., and Collins, M. (2019). Identifying Archaeological Bone via Non-Destructive ZooMS and the Materiality of Symbolic Expression: Examples from Iroquoian Bone Points. Sci. Rep. 9, 11027.

22. Bray, F., Flament, S., Abrams, G., Bonjean, D., Rolando, C., Tokarski, C., and Auguste, P. (2023). Extinct species identification from late middle Pleistocene and earlier Upper Pleistocene bone fragments and tools not recognizable from their osteomorphological study by an enhanced proteomics protocol. Archaeometry 65, 196–212.

23. Coutu, A.N., Taurozzi, A.J., Mackie, M., Jensen, T.Z.T., Collins, M.J., and Sealy, J. (2021). Palaeoproteomics confirm earliest domesticated sheep in southern Africa ca. 2000 BP. Sci. Rep. 11, 6631.

24. Welker, F., Collins, M.J., Thomas, J.A., Wadsley, M., Brace, S., Cappellini, E., Turvey, S.T., Reguero, M., Gelfo, J.N., Kramarz, A., et al. (2015). Ancient proteins resolve the evolutionary history of Darwin's South American ungulates. Nature 522, 81–84.

25. Welker, F., Ramos-Madrigal, J., Gutenbrunner, P., Mackie, M., Tiwary, S., Rakownikow Jersie-Christensen, R., Chiva, C., Dickinson, M.R., Kuhlwilm, M., de Manuel, M., et al. (2020). The dental proteome of Homo antecessor. Nature 580, 235–238.

26. Cappellini, E., Welker, F., Pandolfi, L., Ramos-Madrigal, J., Samodova, D., Rüther, P.L., Fotakis, A.K., Lyon, D., Moreno-Mayar, J.V., Bukhsianidze, M., et al. (2019). Early Pleistocene enamel proteome from Dmanisi resolves Stephanorhinus phylogeny. Nature 574, 103–107.

27. Demarchi, B., Stiller, J., Grealy, A., Mackie, M., Deng, Y., Gilbert, T., Clarke, J., Legendre, L.J., Boano, R., Sicheritz-Pontén, T., et al. (2022). Ancient proteins resolve controversy over the identity of Genyornis eggshell. Proc. Natl. Acad. Sci. USA 119. e2109326119.

28. Hendy, J., Colonese, A.C., Franz, I., Fernandes, R., Fischer, R., Orton, D., Lucquin, A., Spindler, L., Anvari, J., Stroud, E., et al. (2018). Ancient proteins from ceramic vessels at Çatalhöyük West reveal the hidden cuisine of early farmers. Nat. Commun. 9, 4064.

29. Warinner, C., Hendy, J., Speller, C., Cappellini, E., Fischer, R., Trachsel, C., Arneborg, J., Lynnerup, N., Craig, O.E., Swallow, D.M., et al. (2014). Direct evidence of milk consumption from ancient human dental calculus. Sci. Rep. 4, 7104.

30. Scott, A., Power, R.C., Altmann-Wendling, V., Artzy, M., Martin, M.A.S., Eisenmann, S., Hagan, R., Salazar-García, D.C., Salmon, Y., Yegorov, D., et al. (2021). Exotic foods reveal contact between South Asia and the Near East during the second millennium BCE. Proc. Natl. Acad. Sci. USA 118. e2014956117.

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

32. Rüther, P.L., Husic, I.M., Bangsgaard, P., Gregersen, K.M., Pantmann, P., Carvalho, M., Godinho, R.M., Friedl, L., Cascalheira, J., Taurozzi, A.J., et al. (2022). SPIN enables high throughput species identification of archaeological bone by proteomics. Nat. Commun. 13, 2458.

33. Brown, S., Higham, T., Slon, V., Pääbo, S., Meyer, M., Douka, K., Brock, F., Comeskey, D., Procopio, N., Shunkov, M., et al. (2016). Identification of a new hominin bone from Denisova Cave, Siberia using collagen fingerprinting and mitochondrial DNA analysis. Sci. Rep. 6, 23559.

34. Hildonen, S., Halvorsen, T.G., and Reubsaet, L. (2014). Why less is more when generating tryptic peptides in bottom-up proteomics. Proteomics 14, 2031–2041.

35. Zhang, D., Xia, H., Chen, F., Li, B., Slon, V., Cheng, T., Yang, R., Jacobs, Z., Dai, Q., Massilani, D., et al. (2020). Denisovan DNA in Late Pleistocene sediments from Baishiya Karst Cave on the Tibetan Plateau. Science 370, 584–587.

36. Brown, S., Douka, K., Collins, M.J., and Richter, K.K. (2021). On the standardization of ZooMS nomenclature. J. Proteomics 235, 104041.

37. Chen, F., Welker, F., Shen, C.-C., Bailey, S.E., Bergmann, I., Davis, S., Xia, H., Wang, H., Fischer, R., Freidline, S.E., et al. (2019). A late Middle Pleistocene Denisovan mandible from the Tibetan Plateau. Nature 569, 409–412.

38. Svozil, J., and Baerenfaller, K. (2017). A Cautionary Tale on the Inclusion of Variable Posttranslational Modifications in Database-Dependent Searches of Mass Spectrometry Data. In Methods in Enzymology, 586, A.K. Shukla, ed (Academic Press), pp. 433–452.

39. Robinson, N.E., Robinson, Z.W., Robinson, B.R., Robinson, A.L., Robinson, J.A., Robinson, M.L., and Robinson, A.B. (2004). Structure-dependent nonenzymatic deamidation of glutaminyl and asparaginyl pentapeptides. J. Pept. Res. 63, 426–436.

40. Ramsøe, A., van Heekeren, V., Ponce, P., Fischer, R., Barnes, I., Speller, C., and Collins, M.J. (2020). DeamiDATE 1.0: Site-specific deamidation as a tool to assess authenticity of members of ancient proteomes. J. Archaeol. Sci. 115, 105080.

41. Mylopotamitaki, D., Harking, F.S., Taurozzi, A.J., Fagernäs, Z., Godinho, R.M., Smith, G.M., Weiss, M., Schüler, T., McPherron, S.P., Meller, H., et al. (2023). Comparing extraction method efficiency for high-throughput palaeoproteomic bone species identification. Sci. Rep. 13, 18345.

42. Tomorrow (2021). ElectricityMaps API.

43. Urai, A.E., and Kelly, C. (2023). Rethinking academia in a time of climate crisis. Elife 12, e84991.

44. Smith, P., Feijao, C., Ang, C., Politi, C., Flanagan, I., Qu, M., and Guthrie, S. (2023). Advancing environmentally sustainable health research. https://cms.wellcome.org/sites/default/files/2023-08/Research_Sustainability_Report_RAND_Europe_August_2023.pdf.

45. Eichhorn, A. (2023). Academia should go beyond carbon accounting and take action on climate. Nat. Rev. Phys. 5, 548.

46. Burke, I.C. (2010). Travel trade-offs for scientists. Science 330, 1476.

47. Urbina, M.A., Watts, A.J.R., and Reardon, E.E. (2015). Labs should cut plastic waste too. Nature *528*, 479.

48. Achten, W.M., Almeida, J., and Muys, B. (2013). Carbon footprint of science: More than flying. Ecol. Indic. *34*, 352–355.

49. Warinner, C., Korzow Richter, K., and Collins, M.J. (2022). Paleoproteomics. Chem. Rev. *122*, 13401–13446.

50. Richter, K.K., Codlin, M.C., Seabrook, M., and Warinner, C. (2022). A primer for ZooMS applications in archaeology. Proc. Natl. Acad. Sci. USA *119*. e2109323119.

51. Hughes, C.S., Moggridge, S., Müller, T., Sorensen, P.H., Morin, G.B., and Krijgsveld, J. (2019). Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. Nat. Protoc. *14*, 68–85.

52. Proc, J.L., Kuzyk, M.A., Hardie, D.B., Yang, J., Smith, D.S., Jackson, A.M., Parker, C.E., and Borchers, C.H. (2010). A quantitative study of the effects of chaotropic agents, surfactants, and solvents on the digestion efficiency of human plasma proteins by trypsin. J. Proteome Res. *9*, 5422–5437.

53. Lesur, A., Varesio, E., and Hopfgartner, G. (2010). Accelerated tryptic digestion for the analysis of biopharmaceutical monoclonal antibodies in plasma by liquid chromatography with tandem mass spectrometric detection. J. Chromatogr. A *1217*, 57–64.

54. Mackie, M., Rüther, P., Samodova, D., Di Gianvincenzo, F., Granzotto, C., Lyon, D., Peggie, D.A., Howard, H., Harrison, L., Jensen, L.J., et al. (2018). Palaeoproteomic Profiling of Conservation Layers on a 14th Century Italian Wall Painting. Angew. Chem. *57*, 7369–7374.

55. Cappellini, E., Prohaska, A., Racimo, F., Welker, F., Pedersen, M.W., Allentoft, M.E., de Barros Damgaard, P., Gutenbrunner, P., Dunne, J., Hammann, S., et al. (2018). Ancient Biomolecules and Evolutionary Inference. Annu. Rev. Biochem. *87*, 1029–1060.

56. Jensen, T.Z.T., Yeomans, L., Meillour, L.L., Wistoft Nielsen, M., Ramsøe, M., Mackie, M., Bangsgaard, P., Kinzel, M., Thuesen, I., Collins, M.J., and Taurozzi, A.J. (2023). Tryps-IN: A streamlined palaeoproteomics workflow enables ZooMS analysis of 10,000-year-old petrous bones from Jordan rift-valley. J. Archaeol. Sci. Rep. *52*, 104238.

57. Marshall-Cook, J., and Farley, M. (2021). Sustainable Science and the Laboratory Efficiency Assessment Framework (LEAF).

58. Farley, M. (2022). How green is your science? The race to make laboratories sustainable. Nat. Rev. Mol. Cell Biol. *23*, 517.

59. Madhusoodanan, J. (2020). What can you due to make your lab greener? Nature *581*, 228–229.

60. Norrgran, J., Williams, T.L., Woolfitt, A.R., Solano, M.I., Pirkle, J.L., and Barr, J.R. (2009). Optimization of digestion parameters for protein quantification. Anal. Biochem. *393*, 48–55.

61. Phillips, A.S., Szarka, S., and Wheller, R. (2023). Single-day protein LC–MS bioanalysis: can next-generation trypsins cut it? Bioanalysis *15*, 391–405.

62. Strohalm, M., Kavan, D., Novák, P., Volný, M., and Havlícek, V. (2010). mMass 3: a cross-platform software environment for precise analysis of mass spectrometric data. Anal. Chem. *82*, 4648–4651.

63. Tyanova, S., Temu, T., and Cox, J. (2016). The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. Nat. Protoc. *11*, 2301–2319.

64. Andreaki, V., Barceló, J.A., Antolín, F., Gassmann, P., Hajdas, I., López-Bultó, O., Martínez-Grau, H., Morera, N., Palomo, A., Piqué, R., et al. (2022). Absolute Chronology at the Waterlogged Site of la Draga (lake Banyoles, NE Iberia): Bayesian Chronological Models Integrating Tree-Ring Measurement, Radiocarbon Dates and Micro-Stratigraphical Data. Radiocarbon *64*, 907–948.

65. Palomo, A., Piqué, R., and Terradas, X. (2017). La revolució neolítica: La Draga, el poblat dels prodigis.

66. Saña, M., Bogdanovic, I., and Navarrete, V. (2014). Taphonomic evaluation of the degree of historical representation of the archaeological bone samples in anaerobic versus aerobic environments: The Neolithic site of La Draga (Banyoles, Spain). Quat. Int. *330*, 72–87.

67. Welker, F., Hajdinjak, M., Talamo, S., Jaouen, K., Dannemann, M., David, F., Julien, M., Meyer, M., Kelso, J., Barnes, I., et al. (2016). Palaeoproteomic evidence identifies archaic hominins associated with the Châtelperronian at the Grotte du Renne. Proc. Natl. Acad. Sci. USA *113*, 11162–11167.

68. Wilson, J., van Doorn, N.L., and Collins, M.J. (2012). Assessing the extent of bone degradation using glutamine deamidation in collagen. Anal. Chem. *84*, 9041–9048.

69. Bekker-Jensen, D.B., Martínez-Val, A., Steigerwald, S., Rüther, P., Fort, K.L., Arrey, T.N., Harder, A., Makarov, A., and Olsen, J.V. (2020). A Compact Quadrupole-Orbitrap Mass Spectrometer with FAIMS Interface Improves Proteome Coverage in Short LC Gradients. Mol. Cell. Proteomics *19*, 716–729.

70. Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. *26*, 1367–1372.

71. R Core Team (2010). R: A Language and Environment for Statistical Computing.

72. Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., et al. (2019). Welcome to the tidyverse. J. Open Source Softw. *4*, 1686.

73. Charif, D., and Lobry, J. (2007). Seqinr 1.0-2: A Contributed Package To The R Project For Statistical Computing Devoted To Biological Sequences Retrieval And Analysis. In Structural approaches to sequence evolution: Molecules, networks, populations, U. Bastolla, M. Porto, H. Roman, and M. Vendruscolo, eds. (Springer Verlag), pp. 207–232.

74. Kassambara, A. (2018). ggpubr: 'ggplot2' Based Publication Ready Plots.

75. Kay, M. (2023). ggdist: Visualizations of Distributions and Uncertainty in the Grammar of Graphics. IEEE Trans. Visual. Comput. Graph. *30*, 414–424.

76. Dowle, M., Srinivasan, A., Gorecki, J., Chirico, M., Stetsenko, P., Short, T., Lianoglou, S., Antonyan, E., Bonsch, M., Parsonage, H., et al. (2019). Package 'data. table' (Extension of 'data. frame'), p. 596.

77. Xiao, N. (2023). ggsci: Scientific Journal and Sci-Fi Themed Color Palettes for 'ggplot2'.

78. Bengtsson, H. (2020). A unifying framework for parallel and distributed processing in R using futures. Preprint at arXiv. https://doi.org/10.32614/RJ-2021-048.

79. Lucas, A., Scholz, I., Boehme, R., Jasson, S., and Maechler, M. (2023). gmp: Multiple Precision Arithmetic.

80. Wickham, H. (2007). Reshaping Data with the reshape Package. J. Stat. Softw. *21*, 1–20.

81. Gagolewski, M. (2022). stringi: Fast and portable character string processing in R. J. Stat. Softw. *103*, 1–59.

82. Gibb, S., and Strimmer, K. (2012). MALDIquant: a versatile R package for the analysis of mass spectrometry data. Bioinformatics *28*, 2270–2271.

83. Gibb, S., and Franceschi, P. (2023). MALDIquantForeign: Import/Export Routines for 'MALDIquant'.

84. Firke, S. (2023). Janitor: Simple Tools for Examining and Cleaning Dirty Data.

85. Ram, K., and Wickham, H. (2018). wesanderson: A Wes Anderson Palette Generator.

86. Kuznetsova, A., Brockhoff, P.B., and Christensen, R.H.B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. J. Stat. Softw. *82*, 1–26.

87. Bates, D., Mächler, M., Bolker, B., and Walker, S. (2014). Fitting linear mixed-effects models using Lme4. Preprint at arXiv. https://doi.org/10.18637/jss.v067.i01.

88. Venables, W.N., and Ripley, B.D. (2013). Modern Applied Statistics with S-PLUS (Springer Science & Business Media).

89. Box, G.E.P., Hunter, W.G., and Hunter, J.S. (1978). Statistics for Experimenters (Wiley).

90. Fox, J., Weisberg, S., and Price, B. (2022). carData: Companion to Applied Regression Data Sets.

91. Fox, J., and Weisberg, S. (2019). An R Companion to Applied Regression, Third Edition (Sage).

92. Kassambara, A. (2023). Pipe-Friendly Framework for Basic Statistical Tests. R package rstatix version 0.7.2.

93. Shapiro, S.S., and Wilk, M.B. (1965). An analysis of variance test for normality (complete samples). Biometrika *52*, 591–611.

94. Levene, H. (1960). Robust tests for equality of variances. In Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling, I. Olkin and H. Hotelling, eds. (Stanford University Press), pp. 278–292.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
La Draga bone specimens	Autonomous University of Barcelona	8; 2; 21/CGG_1_025259; 36;42; 69/CGG_1_025265
Baishiya Karst Cave bone specimens	Lanzhou University	B437; B442; B904; B912; B1009; B1122
Chemicals, peptides, and recombinant proteins		
Hydrochloric acid	Sigma Aldrich	Cat#258148
Ammonium bicarbonate	Sigma Aldrich	Cat#A6141
Trypsin	Promega	Cat#V5111
Trifluoroacetic acid	Sigma Aldrich	Cat#302031
α-Cyano-4-hydroxycinnamic acid matrix solution	Sigma Aldrich	Cat#70990
Proteomix Peptide calibration mix4	LaserBioLabs	C104
Acetonitrile	Sigma Aldrich	Cat#271004
Formic Acid	Sigma Aldrich	Cat#F0507
Deposited data		
Raw MALDI spectra (.msd format) and code used for generating processed data	This paper	Zenodo: https://doi.org/10.5281/zenodo.8290650
Raw LC-MS/MS data and processed data associated	This paper	ProteomeXchange identifier: PXD045027
Original code used for data processing and visualization	This paper	Dryad: https://doi.org/10.5061/dryad.cz8w9gj8j
Software and algorithms		
mMass	Strohalm et al. 2008[62]	https://doi.org/10.1002/rcm.3444
MaxQuant	Tyanova et al. 2016[63]	https://doi.org/10.1038/nprot.2016.136
R	CRAN	https://cran.r-project.org
Other		
Power Monitor for electricity measurement	Cowell	PMB01
5800 MALDI-ToF spectrometer	AB Sciex	01701
Evosep One	Evosep	N/A
Exploris 480	Thermo Fisher Scientific	N/A

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Dr. F. Welker, Associate Prof., Globe Institute, University of Copenhagen, Denmark, (frido.welker@sund.ku.dk).

Materials availability

Archaeological specimens used for this study have been returned to their curatorial facilities.

Data and code availability

Raw and processed LC-MS/MS data have been deposited on ProteomeXchange via the PRIDE partner repository and are available as of date of publication. Identifier of the dataset is listed in the key resources table. MALDI-ToF MS data used for the ZooMS analysis, along with the code used for merging replicates into single spectra used for manual visual inspection is listed in the key resources table.

All original code has been deposited on Zenodo and Dryad and is publicly available as of the date of publication. DOIs are listed in the key resources table.

Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

METHOD DETAILS

Sample selection

To ensure our observations are consistent among preservation conditions and chronological ages, we sampled bone specimens from two archaeological sites, one Holocene and one Middle Pleistocene, and a total of three different preservational environments, a cave environment, a phreatic environment, and a terrestrial environment. Furthermore, we sampled a single bone tissue, cortical bone, to avoid any possible influence of sampling different bone tissues. Finally, we sampled specimens from a restricted taxonomic diversity to simplify subsequent data analysis, especially for the shotgun proteomics. Finally, although our MALDI-ToF MS analysis is conducted on a comparatively small number of bone specimens, 12, this allowed us to analyze all the exact same extracts using shotgun proteomic methods, providing deeper insights into proteome composition and peptide and amino acid modifications, that would not be observable using MALDI-ToF MS data alone.

Baishiya Karst Cave

Baishiya Karst Cave is a Paleolithic site located in the northeast Tibetan Plateau, in China (35.449°N, 102.571°E, 3280 m a.s.l.). It is currently the only known site in East Asia with both Denisovan fossil remains[37] and sedimentary Denisovan mitochondrial DNA.[35] The six bone specimens included in this paper were collected from Layer 10 of two connected excavation units (T2 and T3). These specimens had, prior to selection, been taxonomically identified to either *Bos* sp. or *Bison* sp. through ZooMS. According to a previous chronological study,[35] the age range of Layer 10 of unit T2 is from 224.8 to 109k years before present (BP), and the specimens thus most likely belong to the late Middle Pleistocene.

La Draga

La Draga is an Early Neolithic (7250-6750 years cal. BP) open air site located in Banyoles, Spain (42.125°N, 2.759°E, 170 m.a.s.l.[64]). The site is known for its remarkable preservation and abundance of objects made of, among others, bone, wood, and marble, as well as basketry, plant material, and other non-worked organic materials.[65] The excellent visual preservation is most likely due to the neighboring lake creating favourable anaerobic conditions for the archaeological site within and adjacent to the lake. This, however, is not the case for the upper levels (which are completely aquatic) and terrestrial inland areas of the site. As a result, the site is divided in a terrestrial area, Sector A, and a phreatic or aquatic preservation area, Sector B, C, and D.[66] We sampled three humeri each from sectors A and B to capture this variation in preservational environments in our study design. All selected specimens were previously morphologically identified as *Bos* sp., with one specimen re-assigned to Cervidae based on our ZooMS and SPIN results.

Bone proteome extraction

Bone powder of approximately 30 mg per specimen was obtained by breaking a bone chip, which was subsequently powdered using a mortar and pestle. Each sample was thereafter split into three different microtubes, of approximately 10 mg each (ProteinLoBind, Eppendorf) for the three digestion time slots of 18, 6, and 3 h (Figure 1). The sample weight was recorded for each sample individually (mean of all specimens = 10.12 ± 3.01 mg standard deviation SD; mean for LD = 12.45 ± 1.58 mg SD; mean for BKC = 7.78 ± 1.86 mg SD). Instruments were cleaned in between each sampling event using 2% Hellmanex III and 70% ethanol, consecutively. Sample extractions were performed in accordance with a set time schedule to prevent any unnecessary freezing and fridge storage of samples after digestion, regardless of the duration of the latter (Figure 1). The following results should therefore not be compounded by differences in fridge or freezer storage time between the three digestion conditions.

Samples were then processed according to the ZooMS protocol after Buckley et al.[31] and Welker et al.[67] Samples were demineralised in 0.6 M hydrochloric acid (HCl) for 24 h. The HCl supernatant was then removed and samples were rinsed thrice in 100 μL ammonium bicarbonate (50 mM, NH_4HCO_3, hereafter AmBic, pH 8.0), for subsequent gelatinisation in a final volume of 100 μL AmBic for 1 h at 65°C. Following gelatinisation, the 100 μL AmBic solution was transferred to a new microtube, to which 0.8 μg trypsin (Promega, #V115A) was added for incubation at 37°C, with mild agitation at 300 rpm (VWR, Thermal Shake *lite*). Digestion occurred for either 3, 6, or 18 h. To stop trypsin digestion, 2 μL of 5% trifluoroacetic acid (TFA) was added to each sample. The digested extracts were then split in two parts for separate analyses via matrix-assisted laser desorption/ionisation-time of flight mass spectrometry (MALDI-ToF MS) and liquid chromatography-tandem mass spectrometry (LC-MS/MS). To assess any potential contamination by non-endogenous peptides, we performed extraction of laboratory blanks alongside the samples for each enzymatic digestion condition.

Mass spectrometry analyses

MALDI-ToF MS and ZooMS data analysis

For ZooMS data analysis, prior to MALDI-ToF MS analysis, peptides were cleaned and desalted using C18 ZipTips (Thermo Fisher) and subsequently spotted in triplicate, consisting of 0.5 μL eluted peptides and 0.5 μL α-cyano-4-hydroxycinnamic acid (CHCA) matrix solution, on a 384-well Opti-ToF MALDI plate insert (AB Sciex, Framingham, MA, 01701, USA) and allowed to air-dry at room temperature. MALDI spectra were automatically acquired with an AB SCIEX 5800 MALDI-ToF spectrometer (Framingham, MA, 01701, USA) in positive reflector mode for MS acquisition. Before sample acquisition, an external plate model calibration was achieved on 13 adjacent MS standard spots with a standard peptide mix (Proteomix Peptide calibration mix4, LaserBioLabs, Sophia Antipolis, France) containing bradykinin fragment 1–5 (573.315 Da), human angiotensin II (1,046.542 Da), neurotensin (1,672.917 Da), ACTH fragment 18–39 (2,464.199 Da) and oxidised insulin B chain

(3,494.651 Da). The concentration in the prepared mixture was between 27 and 167 fmol/µL. The calibration was validated according to the laboratory specifications (resolution above 10,000 for 573 Da, 12,000 for 1,046 Da and 15 to 25,000 for other masses, error tolerance <50 ppm). For the spectra where peptides resulting from trypsin autolysis were detected, an internal recalibration was applied to decrease the error tolerance below 10 ppm (trypsin peptides: 842.509 Da, 1,045.56 Da and 2,211.104 Da). Laser intensity was set at 50% after optimization of signal to noise ratio on several spots, then operated at up to 3,000 shots accumulated per spot, and covering a mass-to-charge range of 1,000 to 3,500 Da for sample analysis. The triplicate data files were merged in R and converted into.msd files. ZooMS taxonomic identifications were assessed using mMass[62] through manual peptide marker mass identification in comparison to a database of peptide marker series for medium-to large-sized mammals.[67] Reference to specific ZooMS peptide markers in this publication is made according to the nomenclature proposed by Brown et al.[36] Glutamine deamidation values were calculated using the Betacalc package.[68]

Shotgun proteomics

For SPIN data analysis, peptide extracts were first separated using an Evosep One (Evosep, Odense, Denmark) with the 100 samples-per-day method (cycle of 14.4 min). Loading of samples was conducted at a flow rate of 2 µL/min using mobile phases of A: 5% acetonitrile and 0.1% formic acid in H_2O and B: 0.1% formic acid in H_2O with a gradient of 11.5 min at 1.5 µL/min. A polymicro flexible fused silica capillary tubing of 150 µm inner diameter and 16 cm long home-pulled was packed with C18 bounded silica particles of 1.9 µm diameter (ReproSil-Pur, C18-AQ, Dr. Maisch, Germany). The column was mounted on an electrospray source with a column oven set at 60°C with the source voltage of +2000 V, along with an ion transfer tube set at 275°C. An Exploris 480 (Thermo Fisher Scientific) was operating in data dependent mode consisting of a first MS1 scan at resolution of 60 000 between *m/z* of 350 and 1400. The twelve most intense monoisotopic precursors were selected if above 2E5 intensity with a charge state between 2 and 6, and were then dynamically excluded after one appearance with their isotopes (±20 ppm) for 20 s.[69] The selected peptides were acquired on an MS2 at an Orbitrap resolving power of 15,000, with the normalised collision energy (HCD) set at 30%, a quadrupole isolation width of 1.3 *m/z* and a first *m/z* of 120. Quality control was assessed on HeLa cells using as QC displayed of 1289 protein groups for 5561 peptides at a repeating sequencing of 2.90% on MaxQuant v.2.2.3.0.[63,70] The following parameters were used for the quality control search: the raw data were searched against the human full proteome, with carbamidomethyl (C) as fixed modification and oxidation (M) and acetyl (protein N term) as variable; digestion was set as tryptic and all other parameters were kept as default.

QUANTIFICATION AND STATISTICAL ANALYSIS

MaxQuant search

All palaeoproteomic.raw files were analyzed using MaxQuant (v.2.3.1) in two different searches. The first search was performed as described in Rüther et al.[32] against the protein sequences database provided there. Download dates are available in the MaxQuant "summary.txt" file available through the ProteomeXchange submission PXD045027. Variable modifications included oxidation (M), deamidation (NQ), Gln (Q) -> pyro-Glu, Glu (E) -> pyro-Glu, and proline (P) hydroxylation. The internal MaxQuant contaminant list was replaced with an in-house database provided by Rüther et al.,[32] Supplementary File "PR200512_HumanCons.fasta"). Since all specimens except for one were identified as belonging to either *Bos* sp. or *Bison* sp., a second search was performed against the whole *Bos taurus* reference proteome (downloaded from Uniprot on 2022-01-20, proteome ID UP000009136 with 23,847 reviewed sequence entries) to explore the presence of other, additional non-collagenous proteins (NCPs). Variable modifications for this search included oxidation (M), deamidation (NQ), and proline (P) hydroxylation. The internal MaxQuant contaminant list was used. Both searches were run in semi-specific Trypsin/P digestion mode. Up to five variable modifications were allowed per peptide and all other settings were left as default for both searches.

Data and statistical analysis

After spectral identification, proteomic data analysis was conducted largely through R v.4.1.2[71] using *tidyverse* v.1.3.1,[72] *seqinr* v.4.2–8,[73] *ggpubr* v.0.4.0,[74] *ggdist* v.3.3.0,[75] *data.table* v.1.14.2,[76] *ggsci* v.2.9,[77] *progressr* v.0.10.0,[78] *gmp* v.0.6–6,[79] *reshape2* v.1.4.4,[80] *stringi* v.1.7.6,[81] *MALDIquant* v.1.21,[82] *MALDIquantForeign* v.0.13,[83] *janitor* v.2.2.0[84] and *wesanderson* v.0.3.6.[85] The R scripts used for the shotgun proteomics analysis are available under Rüther et al.[23] Deamidation was quantified based on spectral intensities, following Mackie et al.[54]

Depending on data types, statistics were calculated using two-way ANOVA (Type II and Type III), linear modeling from lmerTest v.3.1–3,[86] lme4 v.1.1–34,[87] MASS v.7.3–60,[88] and Kruskal Wallis tests[89] from *carData* v.3.0–5,[90] *car* v.3.1–0[91] and *rstatix* v.0.7.2.[92] As prerequisites for ANOVA tests, normal distribution of residuals was checked using the Shapiro-Wilk normality test[93] and homogeneity of the variances was assessed by Levene's test.[94] We used two-way ANOVA (Type II) in this study because we tested two independent variables that presented no significant interactions. When the independent variables showed significant interactions we performed a two-way ANOVA (Type III). F-statistics (F value) are used to determine whether the variance between two normal populations are similar to one another (the mean square of the variable divided by the mean square of each parameter). The p value is the probability of observing a greater absolute value of F-statistics under the null hypothesis, and we follow the standard practice of interpreting $p < 0.05$ as statistically significant.

Electricity consumption and emission intensity

A power monitor (Cowell, model no.: PMB01) was placed in between the heating block (VWR, Thermal Shake *lite*) and the utilised power outlet to measure electricity consumption using either 96-well plates or Eppendorf tubes for 18 h at 37°C. The measurements for both tubes (1.5 mL

Eppendorf Protein LoBind, Eppendorf) and plates (PCR Plate, 96-well, low profile, non-skirted, 0.3 mL, Thermo Fisher Scientific) were separately conducted over the time frame of 18 h, and replicated thrice in total. Measurements started when the heating block had reached a stable temperature of 37°C. The maximum number of tubes, 40 units, were placed in the heating block with 100 μL AmBic in each tube to imitate experiment conditions. Likewise, each well in the 96-well plate was filled with 100 μL AmBic. The emission intensity (gCO$_2$eq; grams of carbon dioxide equivalent) was then calculated by accessing the kWh measured and gCO$_2$eq/kWh values available through Electricity Maps[42] for the dates on which our experiments were conducted. The gCO$_2$eq/kWh values were obtained from various countries (Australia, Brazil, Germany, Denmark, France, Japan, USA, and South Africa). With this selection, we hope to cover a range of countries where high-throughput palaeoproteomic facilities exist. Furthermore, countries differ significantly in the amount of carbon released for each unit of electricity consumed, the so-called carbon intensity, for example due to the use of nuclear energy or largely-completed transitions to wind and solar energy sources. The absolute impact of electricity consumption is therefore very different depending on the country, and our selection of countries aims to also cover this range of carbon intensities. Lastly, emission intensities were calculated for each individual tube and PCR-plate well across the three digestion durations (18 h, 6 h, and 3 h), and for each country included in the study.