Assessing Ancient Protein Studies

Jessica Hendy[1*], Frido Welker[2,5*], Beatrice Demarchi[3,4], Camilla Speller[4], Christina Warinner[6,7,8], Matthew J. Collins[4,5]

[1] Department of Archaeology, Max Planck Institute for the Science of Human History, Jena, Germany.

[2] Department of Human Evolution, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany.:

[3] Department of Life Science and Systems Biology, University of Turin, Italy.

[4] BioArCh, Department of Archaeology, University of York, York, UK.

[5] Natural History Museum of Denmark, University of Copenhagen, Copenhagen, Denmark.

[6] Department of Archaeogenetics, Max Planck Institute for the Science of Human History, Jena, Germany.

[7] Department of Anthropology, University of Oklahoma, Norman, USA.

[8] Institute for Evolutionary Medicine, University of Zürich, Zürich, Switzerland.

*Equal author contribution.

Abstract

Palaeoproteomics is an emerging neologism used to describe the application of mass spectrometry (MS)-based approaches to the study of ancient proteomes. As with palaeogenomics (the study of ancient DNA, aDNA), it intersects evolutionary biology, archaeology and anthropology, with applications ranging from the phylogenetic reconstruction of extinct species to the investigation of past human diets and ancient diseases. However, there is currently no explicit consensus regarding standards for data reporting, data validation measures, or the use of suitable contamination controls in ancient protein studies. Additionally, in contrast to the aDNA community, no consolidated guidelines have been proposed by which researchers, reviewers and

editors can evaluate palaeoproteomics data, in part due to the novelty of the field. Here we present a series of precautions and standards for ancient protein research that can be implemented at each stage of analysis, from sample selection to data interpretation. These guidelines are not intended to impose a narrow or rigid list of authentication criteria, but rather to support good practices in the field and to ensure the generation of robust, reproducible results. As the field grows and methodologies change, so too will best practices. It is therefore essential that researchers continue to provide necessary details on how data were generated and authenticated so that the results can be independently and effectively evaluated. We hope that these proposed standards of practice will help to provide a firm foundation for the establishment of palaeoproteomics as a viable and powerful tool for archaeologists, anthropologists, and evolutionary biologists.

Introduction

The advent of high-sensitivity mass spectrometry in the past two decades has allowed palaeoproteomics to become increasingly relevant in the fields of archaeology and evolutionary biology. Not only can individual proteins from archaeological and palaeontological contexts be studied, but one can also analyse the complex mixtures of proteins produced by individual organisms (proteomes) or groups of organisms (metaproteomes) found within ancient samples[1–3]. This has facilitated the phylogenetic reconstruction of extant and extinct species[2,4–6], including that of hominins[7], the mechanistic investigation of protein degradation pathways[8], studies of diagenetic and *in vivo* protein post-translational modifications (PTMs)[9–11], the reconstruction of human diet and subsistence patterns[3,12], and the characterization of past human diseases[3,13–16]. The range of tissues and substrates that can be analyzed is similarly broad, including bone, antler, dentine and enamel[1,7,17–19], eggshell[8,20], skin and soft tissues[13,14], dental calculus[21], preserved food remains[22–25], potsherds and ceramic vessels [26–28], bindings and glues[28–31], paint binders[32–34], textiles and leather[35,36] parchment[37], mortars[38–40] and soil[41].

While palaeoproteomics is a relatively young discipline, the survival of ancient proteins over archaeological and geological timescales has been studied since Abelson's discovery of amino acids in fossils in 1954[42]. Important studies in organic geochemistry were conducted, for

example, on mechanisms of degradation [43–46], including on the likelihood of finding collagen in dinosaur bones, [47], and on the difference between preservation of soft tissues and of molecular-level information[48]. These studies highlighted that proteins are more resistant to degradation than DNA due to their chemical and physical properties, but that diagenesis nonetheless affects the protein sequences, so that short and altered peptide fragments tend to be recovered from ancient substrates, providing the first challenge for protein identification and authentication. While these early technologies might be outdated, the conclusions that were drawn from them are not, and these initial studies represent the foundation for distinguishing between endogenous (albeit degraded) proteins and contamination.

Palaeoproteomics, as a relatively young discipline, faces many of the same epistemological and analytical challenges that the field of ancient DNA did roughly two decades ago. Recent proteomic studies of ancient proteins exhibit a wide disparity in data reporting standards, protein authentication measures, and procedures taken to avoid protein contamination (Supplementary Table 1). Many of the principles put forward in the field of ancient DNA, such as isolation of work areas, the inclusion of negative controls, and the demonstration of appropriate molecular behaviour, provide a useful starting point, but additional measures are necessary. In particular, the conserved nature of proteins compared to DNA renders the authentication of ancient protein sequences more challenging than that of ancient DNA. For example, within palaeogenomics, the presence of multiple mitochondrial DNA sequences within a single DNA extract can be used to both detect and quantify modern human contamination[49,50]. In contrast, the low amount of intraspecific amino acid sequence variation generally makes it impossible to use protein sequence variation as a criterion by which to detect the presence of multiple contributing individuals of the same species to a single sample. Nevertheless, many concrete steps can be taken in the field, in the laboratory, and during analysis to mitigate the dual challenges posed by contamination and degradation and to improve the identification of endogenous proteins. Authentication criteria are essential because reporting the identification of extraordinary, purportedly ancient proteins without sufficient evidence of authentication can damage the credibility of this emerging research area[51].

Here we collate and suggest a number of "best practices" for the sampling, generation, analysis, and reporting of ancient protein sequence data in the scientific literature (summarized in Box 1), which we hope will be useful to researchers, reviewers and editors. Building on our collective experiences across six different palaeoproteomics laboratories and in our roles as both researchers and reviewers, our suggestions are intended to complement - not replace - previously established guidelines for modern proteomic studies (e.g.[for example 52]). Some aspects might appear common practice, but we state them nevertheless as the emerging research area encompasses scholars in a variety of disciplines (for example, analytical chemistry, evolutionary biology, organic geochemistry, archaeology, paleontology). Criteria that may appear obvious or standard to members of one field may be perceived as innovative or novel to members of another. One aim of this article is therefore to establish a common foundation so that researchers across disciplinary backgrounds can contribute to the growth of this new field. We are aware that as new experimental and data analysis strategies emerge these guidelines will no doubt require further refinement and amendment. However, in the spirit of the influential ancient DNA research principles proposed by Gilbert et al[53], we emphasise that, at a minimum, researchers must provide details on how data were generated and authenticated, so that others may be able to effectively evaluate ancient protein identifications.

Towards a Standardised Practice of Palaeoproteomics

1. Selection and Sampling

In order to publish high-impact studies within a competitive research climate, it can be tempting to apply "fast science" - to focus immediately on exceptional samples of great antiquity, rarity or their perceived importance. However, archaeological and palaeontological samples are irreplaceable, and have high cultural value. Therefore, preceding the inception of research projects, researchers must consider whether destructive sampling is necessary, whether alternative non-destructive methodologies[54–56] could be employed, and/or what techniques can be utilized to reduce the sample size necessary for analysis[57]. Further research is also needed to explore the integration of multiple biomolecular analyses from minimally-sized samples. This is

especially vital as new methodological advances will no doubt led to further insights from samples in the future. Finally, pilot studies should be conducted in order to test the efficiency of the analytical method using less precious materials and/or artificial diagenesis on modern analogues.

In selecting samples for destructive analysis, several considerations should be kept in mind. The choice of samples should be governed by an awareness of the nature and impact of diagenesis - the physico-chemical changes that affect both the organic and inorganic components of a sample through time. Diagenesis is driven by a complex network of reactions, including chemical degradation (e.g., temperature- and age-inducing peptide bond hydrolysis and amino acid racemisation) and molecular breakdown driven by environmental factors during burial and storage (e.g., microbial decomposition, acid decalcification and water fluctuation) [58,59]. In general, this will reduce the concentration of endogenous proteins, alter their sequences (e.g., some amino acids will be deamidated), and provide opportunities for exogenous proteins (contamination) to be incorporated in the sample. Some substrates may harbor better potential for preserving endogenous proteins than others. For example, mineralized samples (such as bone, dental calculus and eggshell) provide a better preservational environment for proteins than other substrates. The presence of a mineral phase can provide protection from degradation driven by external factors, and mineral-organic binding may facilitate the survival of certain peptides by slowing down peptide bond breakdown[8]. There may also be differences in protein preservation among different mineralized substrates. For example, peptides may persist longer in closed systems such as eggshell than open systems such as bone[8,60]. However, despite these broad generalisations, preservation among individual samples is often highly variable, even from within the same archaeological site [61]. Although reduced sample preservation in the form of low peptide abundance poses technical challenges, the increasing sensitivity of mass spectrometers partially mitigates this problem and enables protein identification from very low-abundance peptides [8].

Several steps can be taken prior to palaeoproteomics analysis in order to evaluate protein preservation and to identify potential sources of contamination during burial and storage. One

approach is to assess the elemental composition of samples, as organic nitrogen is a proxy for protein. Pyrolysis-GC/MS and LC (HPLC) can be used to detect the presence of amino acids [62] in any putative proteinaceous sample. This is especially important for very old samples, whereby an absence of amino acids in a given archaeological or paleontological sample have been used to challenge claims for the detection of protein sequences in fossil samples [63]. Concentration and compositional analyses can additionally assess the yield, and in some cases, the character of the preserved proteins[64–66]. For example, because decomposition products are retained in closed systems (e.g., bleached eggshells), the proportion of free amino acids can reveal the extent of diagenetic hydrolysis, and this can be complemented by assessment of amino acid racemization, i.e., the increase in concentration of D-amino acids[8]. In samples containing collagen, peptide mass fingerprinting (also termed ZooMS, for Zooarchaeology by Mass Spectrometry)[17] may also be useful as a screening technique to assess the extent of preservation and of diagenetically-induced glutamine deamidation[67,68] prior to large-scale application of this method at a site or in advance of laborious and more expensive shotgun proteomic analysis. Regardless of the technique used to assess protein preservation and integrity, it is recommended to first analyze a small subset of samples in order to establish feasibility for a proteomics study of a given collection before proceeding to destructive analysis of a larger batch of samples. Such assessment and screening should be reported alongside other downstream measures of authentication and interpretation.

In addition to critical sample choice and pilot screening, at the sampling stage researchers should also be mindful that the application of consolidants, resins and glues may introduce contamination or mass spectrometry interference. The use of chemical stabilizers is widespread in museum conservation practice[69], and such treatment may result in the unintentional introduction of modern proteins, such as animal collagens in glues, plant proteins in natural resins, or insect proteins in shellac. Hence, researchers should be mindful of the post-excavation history of samples and know that records of such treatments are often missing for 19th and early 20th century collections (Figure 1). Additionally, researchers and curators should avoid the use of plastic films, such as parafilm, as these polymers can cause mass spectrometry interference.

Furthermore, material intended for ancient protein analysis should only be handled with non-latex gloves, as skin proteins and latex may introduce additional contamination.

2. Laboratory Considerations

Contamination is a central concern in any palaeoproteomics project as it potentially provides false insights into protein composition, phylogeny, and protein modification. Contamination can be introduced at nearly any stage of burial, excavation, storage and analysis (Figure 1), but a number of concrete measures can be taken to reduce contamination from modern proteins in the laboratory environment, as well as cross-contamination between ancient samples. Such measures should be described in publications and at a minimum include extraction blanks, the wearing of protective clothing including non-latex gloves (latex is a natural product, containing proteins), the use of clean surfaces and equipment (e.g., washed with bleach solution, 70% ethanol, or baked glassware), chemically pure reagents, and no reuse of consumables.

Laboratories analysing ancient proteins should make attempts to reduce the presence of proteinaceous material in the background laboratory environment, including keratins from wool, hair, and skin, as well as common protein-based laboratory reagents[70]. Steps to achieve this may include wearing synthetic or cotton clothing (no wool, silk, rubber, or leather), covering exposed skin on the hands and arms at all times, and using facemasks and hairnets. Additionally, protein-based laboratory reagents, such as bovine serum albumin (BSA) and chicken lysozyme, should be avoided. If available, the use of a dead air box or positive pressure laminar flow hood is also encouraged in order to provide a sterile or semi-sterile environment where ancient samples can be handled safely.

Cross-contamination from modern proteins can be minimized by separating as much as possible the initial stages of ancient protein research (sampling, extraction, and protein digestion) from other laboratories or environments where modern proteins are handled, as is common in the field of ancient DNA. The extraction and digestion of ancient proteins should be performed in a location separate from experiments working with modern material (e.g., modern food products, cell cultures or tissue studies). In the absence of full separation, spurious contamination events can still occur even if precautions are undertaken to reduce cross-contamination, contributing to

doubt when unexpected or extraordinary findings are observed. For example, a recent study reported the identification of two Crimean-Congo hemorrhagic fever virus (CCHFV) peptides in five of six early Iron Age (750-400 BCE) mortuary vessels from Germany [71]. Today, the distribution of this tick-borne virus is limited to the Balkans and parts of Asia and Africa, and little is known about its origins or history, hence making its incidental discovery in Iron Age Germany an extraordinary finding. However, it cannot be overlooked that the research was performed at the University of Texas Medical Branch in Galveston, Texas, a world leader in the study of viral pathogenesis (including CCHFV), nor that the two CCHFV peptides identified are also components of synthetic vectors (reverse genetics vectors pT7-M and pT7-M-ASKA) used to study viral virulence[72]. Hence, to avoid instances of cross contamination, as well as lingering doubts over possible cross-contamination events, we advocate the use of dedicated extraction environments for ancient proteins.

Cross contamination from *ancient* proteins, as opposed to modern, should also be minimized through cleaning of sample processing areas and equipment, by avoiding the reuse of consumables, and by preparing fresh reagents for each set of sample extractions. Care should also be taken when opening sample tubes to avoid splashing, dripping or aerosol formation, and samples should not be crowded into tube racks or centrifuges, but rather spaced out with one or more empty wells between samples.

In order to characterize and monitor background laboratory contamination (including the presence of potential contaminants in reagents or consumables), blank extractions should be performed alongside extractions, and this data should be analyzed, reported and made available in a similar manner to the ancient samples under investigation. This applies to both small-scale experiments on highly valuable samples, as well as to large-scale studies involving hundreds to thousands of samples, such as ZooMS collagen peptide mass fingerprinting of ancient bone fragments [17].

We note that several ancient protein studies report the use of chemical pre-treatments to remove potential surface contamination prior to protein extraction (including ammonium-bicarbonate[73], EDTA[74], or bleaching[60,75,76]). Such steps have proven moderately

successful in ancient DNA studies[77–79], but to our knowledge these techniques have not been rigorously tested on ancient protein samples, with the exception of bleaching on carbonate substrates. Research on the effectiveness of protein decontamination techniques on different sample substrates is much needed. For example, although mechanical surface removal may be effective for some sample types, bone is highly porous and if the sample has been exposed to phases of wetting, or even significant changes of humidity, there is the potential that surface contaminants have migrated below the surface. Additionally, although strong chemical oxidants are potentially useful for removing both surface and subsurface contaminants, they also have the potential to damage surviving endogenous proteins as well, unless the ancient proteins are protected within the intra-crystalline fraction of the mineral matrix[75,76,80].

A number of protocols have been reported for the extraction of ancient proteins, particularly for the extraction of bone protein [81–83], and include protocols based on SDS buffers and polyacrylamide gels ([24,84]), Filter-Aided Sample Preparation (FASP)[1,85,86], and Gel-Aided Sample Preparation (GASP)[87,88]. However, the efficacy of these protocols, their downstream effect on protein identification and resulting chemically-induced modification have not been systematically compared in studies of ancient proteins, although examples exist that compare their performance on modern material [89,90].

3. Mass Spectrometry

The current generation of mass spectrometers are powerful high-performance instruments, and the hardware and operational costs of such systems typically exceed the budget of individual labs. Consequently, most palaeoproteomics research projects utilize mass spectrometers at institutional core facilities, such as those available at many universities, medical schools, and hospitals. In keeping with standards for modern proteomic analyses, instrument parameters such as the LC column type, MS/MS model, and collision cell type should be described in the manuscript, even when ancient protein extractions are run at an external core facility[91]. These core facilities typically analyze hundreds of samples (LC-MS/MS) to thousands of spots (MALDI-TOF-MS) per year on a single instrument. Because of this, instrument carryover (i.e., the delayed elution of peptides from previous LC runs using the same HPLC column) is a serious

concern, as clients may have little control over how frequently the instrument is cleaned, how often the HPLC columns are changed, or which samples are analyzed before an ancient protein study. It is important that palaeoproteomics researchers discuss their requirements directly with the researchers and/or technicians in charge of analysing their samples. Palaeoproteomics projects must also build controls into their own research design in order to detect and mitigate potential cross-project and cross-sample carryover events.

Injection blanks or wash buffers should be run before and between each sample during LC-MS/MS analysis in order to clean the column and identify peptide carryover, as peptides persisting in LC columns have the potential to contaminate subsequent protein injections during an MS/MS run (Figure 2). The results of these injection blanks (which are distinct from extraction blanks) should be reported in publications, with semi-quantitative analyses of the data (see Demarchi et al.[8]; Figure 4). Researchers may need to investigate the extent of carryover in their mass spectrometry set-up before proceeding with sample loading and analysis. In particular, peptides that display strong binding affinities to mineral phases in archaeological/palaeontological material and thus persist through time, may also be those peptides that adhere to LC columns. Therefore, carryover may particularly impact those peptides that we wish to characterise, and thus monitoring the presence of peptides in injection blanks is vital. After flushing the system prior to beginning a palaeoproteomics run, it is recommended to to inject old, very precious samples first but otherwise randomize the order of the samples in order to avoid batch effects, but to record the sample order so that any suspicious data patterns (e.g., very old samples testing positive for a given protein only if they are run immediately after a very young sample) that may arise can be identified and the samples reanalyzed, if necessary. Details on the injection device and LC columns (such as those relating to the autosampler loop, flow rate and cartridge systems) can be provided in manuscripts.

Replication is optimal for validating results, in particular for critical samples or for extremely novel results[8]. There are several strategies for validating through replication, including experimental replication through the complete re-extraction of the same sample in the same laboratory (or, more optimally, in an independent laboratory), or an analytical replication

through repeated MS/MS analyses of the same protein extract. We recognize that in cases of small amounts of starting material or very rare or precious specimens, it may not be possible to perform multiple experimental replications. We also realize that replication in independent laboratories might place a significant burden on newly establishing research groups due to the high cost of the analyses and the relatively small number of laboratories currently specializing in ancient protein analysis. Nevertheless, independent replication is a powerful method of validation that should be performed, if at all possible, when reporting novel, extraordinary or unexpected findings. However, it should be noted that in both cases any contestimated peptide and protein false discovery rates (FDR) amination occurring prior to the introduction of a sample into an ancient protein laboratory will not be identified or resolved by replication (Figure 1), reiterating the need for care during sample selection.

4. Peptide and Protein Identification

Once mass spectra have been generated, their interpretation will lead to the identification of peptides sequences and, subsequently, of proteins. As with modern studies[92] , at a minimum, essential information should be provided on search tolerances (both MS1 and MS2), fixed and variable protein modifications, peptide-spectrum matches (PSM) score cut-offs, peptide e-values, whether *de novo* and/or error-tolerant matches were allowed, and which algorithm was used to conduct these searches (e.g. Mascot, Sequest). In keeping with modern protein studies, protein identifications should be made on the basis of a minimum of two supporting peptides, and should be reported. Protein sequence databases should be accessible, either as supplementary information or by clear directions to online repositories (e.g., UniProt, including the date at which a repository was accessed or downloaded). All novel amino acid sequences should be supported by more than one MS/MS peptide-spectrum match (PSM)[4]. Where possible, manual *de novo* verification should be used as a support for novel amino acid sequences[8].

Spectral analysis should allow for the types of diagenetic protein modifications typically encountered when dealing with archaeological and palaeontological material, such as glutamine and asparagine deamidation, possibly methionine and tryptophan (di-)oxidation, the formation of pyroglutamic acid, as well as peptide cleavages unrelated to experimentally-derived enzymatic

digestion. However the increased dynamic range of instruments mean that low abundance peptides from non-standard tryptic cleavage[93] and variations in both commercial trypsin performance[94] and in-source fragmentation[95] may be mistaken for hydrolytic damage. Given the high error rate (both false-negatives and false-positives) associated with *de novo* and error-tolerant search engines, researchers should substantiate claims based on the identification of novel protein sequences. This can be achieved using statistical parameters requiring near-complete fragment ion series in multiple spectra overlapping the position of interest, and/or actualistic bioinformatic experiments where the correct sequence is known but removed from the searched database before analyzing ancient samples using a similar bioinformatic workflow when no modern reference sequences are available[4]. Validation of novel peptide sequences can be achieved by incorporating such modified amino acid sequences into a second round of analysis with a modified sequence database[4,7].

Additionally, researchers should carefully consider their selection of reference databases during data analysis, and should always include microbial and/or common contaminant reference databases as appropriate. The failure to select appropriate databases may result in peptide misassignment or even protein misidentification, and taxonomic misassignment is an especially common problem when using small, curated databases. For example, Swiss-Prot, a manually annotated and non-redundant protein sequence database of reviewed protein sequences, contains the nearly complete proteomes of many model organisms, such as mouse (*Mus musculus*) and human (*Homo sapiens*), but only partial proteomes of other taxa, such as sheep (*Ovis aries*), goat (*Capra hircus*), cow (*Bos taurus*), and pig (*Sus scrofa*). Eukaryotic peptide searches against Swiss-Prot tend to result in accurate *protein* assignments, however incorrect *species* identification may occur due to protein homology and when non-model organisms are absent or underrepresented in the database. For example, in a recent analysis of proteins extracted from a medieval sheep tooth using Swiss-Prot as the search database, it was found that only 20% of the identified eukaryotic proteins were assigned to sheep, while the remaining proteins were misassigned to cattle, human, mouse, pig, and goat [3]. In each case, the incorrect species was assigned when the relevant sheep protein was absent from the Swiss-Prot database (Supplementary Table 2). Such database bias is obvious when analyzing archaeological tissues

that originate from a single animal, but it poses more serious problems when analyzing metaproteomes, such as those extracted from ceramic residues or dental calculus. Here, multiple species might be expected from a single sample, and database bias must be accounted for in order to avoid the reporting of analytical artifacts and "phantom" taxa.

Because handling of archaeological and palaeontological specimens during excavation and curation provides plenty of opportunities for human or animal protein contamination or cross-contamination from other artefacts (Figure 1), it is recommended to include possible human contaminating proteins in reference databases in searches of non-human tissues (for example, animal bones). Ideally this also includes human collagen type I sequences, given this particular protein's resilience to degradation and its presence in the dermis of the skin. Additionally, other skin proteins such as desmoglein-1 (DSG1), dermcidin (DCD), junctional plakoglobin (JUP), and of course keratins (both from humans and animals) are recurring contaminants. Contaminating keratins may derive from skin and clothing, but also potentially from brushes or other equipment used in sample preparation and conservation. Future studies focusing on the analysis of mummified skin, ancient furs and textiles will need to address the problem of how to reliably distinguish ancient from modern skin proteins (e.g., through the study of diagenetic protein modifications). Supplementary Data 1 contains a list of commonly encountered contaminants in proteomics laboratories, including the common Repository of Adventitious Proteins (cRAP)[96]. Additional lists containing common background contaminants can be found in the Contaminant Repository for Affinity Purification (CRAPome)[97]. One should keep in mind that some of the proteins in Supplementary Data 1 may represent endogenous proteins depending on the type of sample analyzed (e.g., keratins in furs, egg white proteins in paintings, or albumin in bone).

5. Data Interpretation and Authentication

Following data generation, several additional analyses can be performed to further authenticate and affirm the validity of the results. Like DNA, proteins undergo predictable forms of diagenetic alteration over time, so much so that there is an established field of amino acid / protein diagenesis geochronometry[98], and documentation of diagenetic changes in ancient

samples has been suggested as a useful authentication tool. In particular, diagenetically-induced modifications such as glutamine and asparagine deamidation and the presence of non-enzymatic cleavages of individual proteins are expected to occur in ancient samples[3,7,8,12,15,16]. Some studies have aimed to contrast such diagenetically-derived protein modifications between different proteins identified in the same sample[7,99], allowing the potential separation of endogenous human proteins from contaminating human proteins.

Researchers should also be mindful that amino acid modifications can result in an amino acid position having a total mass equaling that of another amino acid. For example, in the case of the whey protein beta-lactoglobulin reported in Warinner et al.[12], it was observed that one of the protein variant sites that distinguishes Bovinae (cattle, yak, and buffalo) from Caprinae (sheep and goats) is an amino acid residue that is aspartic acid in Bovinae, asparagine in sheep, and lysine in goats (Figure 3a). However, the deamidation of asparagine results in its conversion to aspartic acid (Figure 3b) and hence it is not possible to distinguish an unmodified Bovinae residue (D) from a deamidated sheep residue (de. N) at this position (Figure 3c). Only the identification of an unmodified asparagine (N) or a lysine (K) would therefore allow species discrimination at this site in most situations[7,12]. The presence of diagenetic modifications is particularly challenging for older samples, where deamidation might have converted all surviving endogenous asparagines andr glutamines to aspartic acid and glutamic acid respectively, an issue encountered recently for a Middle Pleistocene rhinoceros proteome [99]. Another example of sequence ambiguity relevant to the most common bone protein, collagen, is the incomplete fragmentation of a proline-serine peptide bond, which produces a peptide fragment ion isobaric to hydroxyproline-alanine. Cleavage N-terminal to Pro ('the proline effect') is enhanced whilst cleavage C-terminal to proline in MS^2 is depressed[100]. Proline hydroxylation is the most common post-translational modification of collagen, and Ser/Ala is one of the most common substitution pairs; therefore differentiating serine (in effect *hydroxy*alanine) from alanine C-terminal to (hydoxy)proline is especially difficult[10,101].

6. Data Integration, Data Sharing and Review

Combining proteomic approaches with other biomolecular techniques, where possible, is encouraged, as multiple approaches can be used to supplement or support novel proteomic findings. For example, ancient mtDNA sequences have been used to support palaeoproteomic analyses of hominin taxonomy [7], lipid and proteomic approaches have been used in combination to detect early Bronze Age cereal grains[88], and proteomic and isotopic approaches have been used together to identify ancient milk consumption[12].

In the era of 'big data' many research communities are mandating the long-term curation of raw datasets in a publicly accessible form, and an updated list of community-recognised repositories is maintained by the journal *Scientific Data*[102]. The sharing of raw and processed mass spectrometric data in public repositories such as the ProteomeXchange[103] is therefore considered a necessity. Accessing and reanalysing raw data is one way that other researchers can test a study's bioinformatic workflow in their own environment. Additionally, archiving allows data to be re-searched in future analyses, and may lead to the identification of additional proteins as reference sequence databases are updated and expanded. This is especially relevant for valuable cultural heritage and human/hominin remains, which might not be available for subsequent re-extraction and destructive analysis. Finally, the public sharing of ancient protein data allows such data to be integrated with future biomolecular analysis using different or similar methods, and more generally *"help[s] build rigorous and reliable scientific practices even in the presence of complex experimental challenges*" [104].

Fundamentally, we call for a critical approach towards the validation of results and data presented in ancient proteins studies, and stress that "checking all the tick boxes" does not automatically validate a study. Following Gilbert et al.[53], we suggest that reviewers and editors consider whether the following questions are sufficiently addressed: 1) Are sufficient measures taken to minimize contamination in the laboratory, and do data analysis strategies take potential contamination and degradation into consideration?; 2) Is adequate proof of authentic, ancient protein identification presented?; and 3) Is sufficient information presented for independent bioinformatic replication and can the resulting data be examined? We also suggest that editors include modern proteomics experts when evaluating ancient protein studies, particularly for

ancient protein studies employing novel technical and methodological tools, in addition to experts in the archaeological or palaeontological context of the research.

Perspective

Palaeoproteomics holds enormous potential to dramatically expand archaeological, palaeontological and evolutionary research. In light of this promise, we have raised key considerations and have recommended standards for the generation and reporting of ancient protein data with the view that these suggestions will aid non-specialist readers and reviewers of ancient protein publications, as well as assist researchers improve palaeoproteomic study designs. Undoubtedly, with the emergence of new experimental and bioinformatic strategies for characterizing protein degradation and contamination, as well as improved tools for protein validation and authentication, these guidelines will require debate within the community, as well as further refinement and updating. However, it is our hope that the standards of practice presented here will help to provide a firm foundation for the consolidation of palaeoproteomics as a robust tool for evolutionary biology, anthropology and archaeology.

Competing interests

The authors declare no competing financial interests.

Contributions

J.H. and F.W. conceived of the manuscript. All authors wrote and contributed to the main text.

Correspondence

Correspondance should be addressed to Jessica Hendy (ORCID: 0000-0002-3718-1058) hendy@shh.mpg.de or Frido Welker (ORCID: 0000-0002-4846-6104) frido_welker@eva.mpg.de

Acknowledgements

We thank Enrico Cappellini for valuable comments on a previous version of the manuscript. We are grateful to Kirsty Penkman, Jane Thomas-Oates, Julie Wilson, and Roman Fischer, as well as Christian Trachsel and Jonas Grossmann (Functional Genomics Center Zürich) for support and helpful discussions. This research was supported by the Max Planck Society (J.H., F.W., C.W.) and its Donation Award (to J.H., C.W.), the US National Science Foundation BCS-1516633, BCS-1523264, and BCS-1643318 (to C.W.), the European Research Council under the European Union's Horizon 2020 research and innovation program under grant agreement numbers STG-678901-FOODTRANSFORMS and STG-677576-HARVEST, the "Rita Levi-Montalcini Young Researchers Programme" (to B.D.), the Wellcome Trust [grant no 108375/Z/15/Z], the Leverhulme Trust [Philip Leverhulme Prize] (to C.S.), the VILLUM FONDEN [grant no 17649], a Danish National Research Foundation Niels Bohr Professorship and ERC Investigator Grant 295729-CodeX (M.J.C) and the US National Institutes of Health R01GM089886 (to C.W.).

Supplementary Data 1. FASTA formatted file containing proteins (in)frequently identified as likely contaminants in standard palaeoproteomic research. This contains commonly encountered contaminants in proteomics laboratories, and consists of the common Repository of Adventitious Proteins, with the addition of several protein sequences we sometimes encounter as contaminants in our experiments.

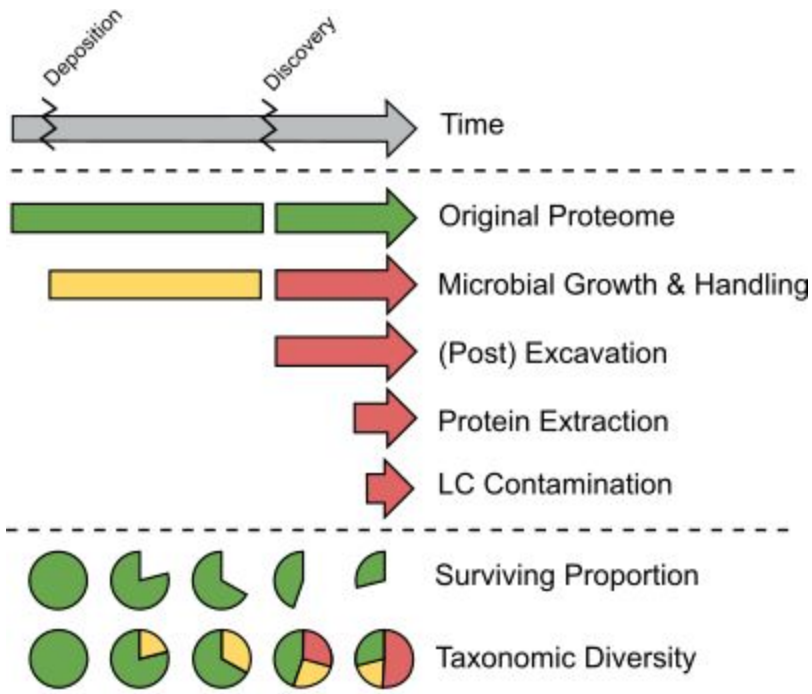


Figure 1. Schematic depiction of ancient proteome compositional changes through time. Initially, the proteome is solely composed of endogenous proteins (green), which may already represent a mixture of taxonomic origins in cases of microbiome samples, food residues, or infected tissues. After deposition, substrates will be rapidly colonized by bacteria and fungi (yellow), some of which might be of interest in future studies. During excavation, curation, and storage, additional contamination can occur, primarily due to human handling and through protein-based consolidants (for example human keratins or animal-based glues; in red). A definitive source of contamination is introduced during sample preparation through the deliberate addition of trypsin, or another protease. Laboratory cross-contamination from both modern and ancient sources can occur during both extraction and LC-MS/MS analysis. Throughout the scheme, proteome complexity and protein concentration of the endogenous proteome decrease. Conversely, there is an increase in the proportion of contaminating proteins, both of vertebrate and non-vertebrate origin. Time not to scale. Proportions are used to illustrate general developments and do not necessarily reflect observed frequencies. Modified from Welker[105].

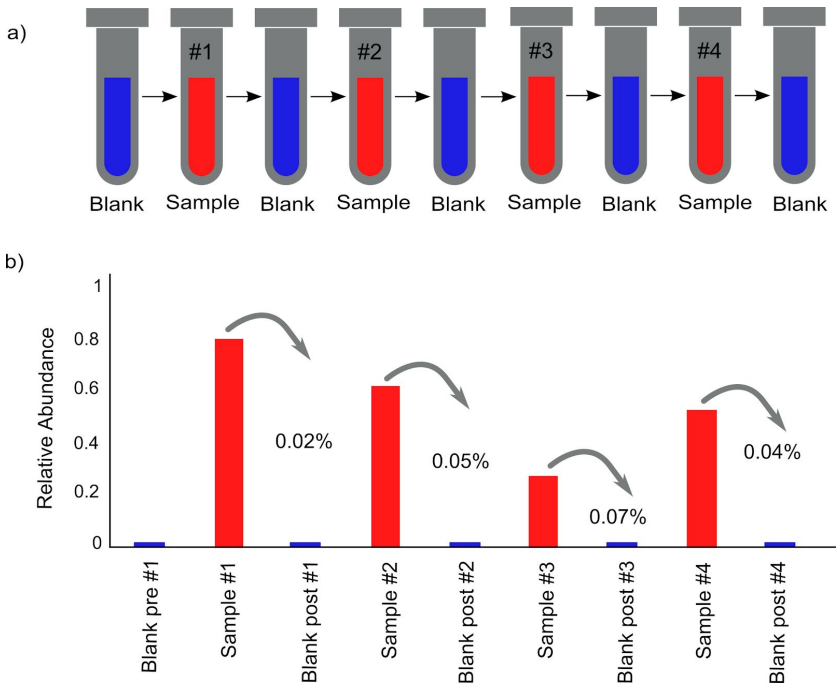


Figure 2. Injection blanks in LC-MS/MS. (**a**) Each sample is preceded and followed by at least one injection blank within the LC column, which (**b**) allows the assessment of peptide carryover between different experiments and samples (following Demarchi et al.[8]). Within this scheme, the extraction blank is analyzed as if representing one of the samples.

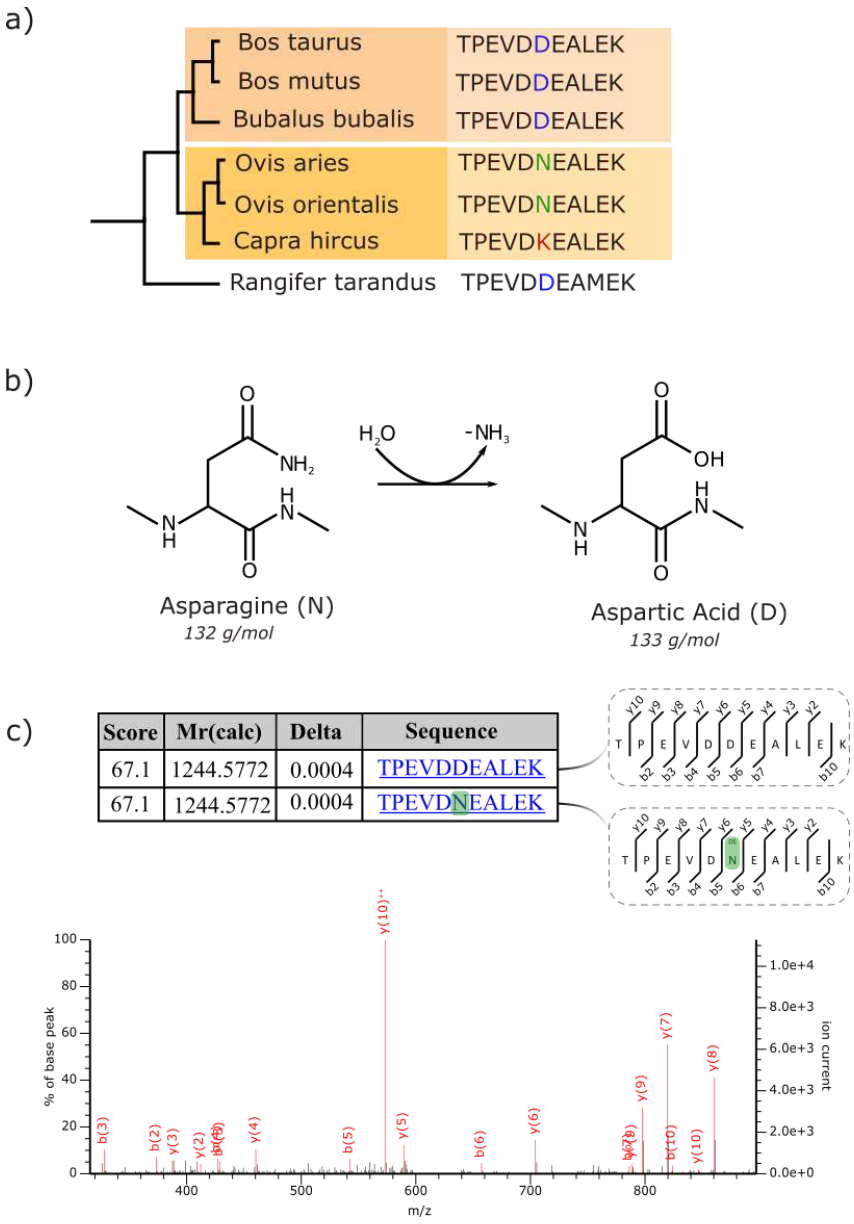
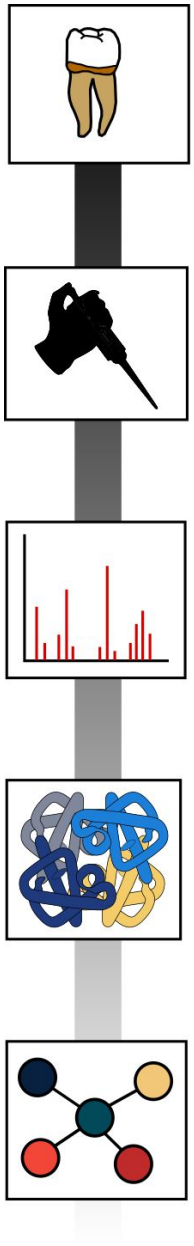


Figure 3. Damage-induced sequence ambiguity affects peptide taxonomic assignment for the whey protein beta-lactoglobulin. (**a**) An important variant site that distinguishes Bovinae (cattle, yak, and buffalo) from Caprinae (sheep and goats) is an amino acid residue that is aspartic acid (D) in Bovinae, asparagine (N) in sheep, and lysine (K) in goats. However, the deamidation of asparagine results in its conversion to aspartic acid (**b**) and Mascot protein identification software is unable to distinguish an unmodified Bovinae residue (D) from a deamidated sheep residue (de. N) at this position (**c**). Data from [12]. Incorrect precursor selection during MS analysis can result in similar ambiguity by selecting isotope-containing peaks for fragmentation, instead of the monoisotopic peak.



Sample Selection

Keep the sample size of destructive analysis to a minimum

Be mindful of the post-excavation history of samples (e.g. treatment with glues)

Perform a screening technique or pilot study to minimize unnecessary sample destruction

Laboratory Considerations

Use sterilized equipment and pure reagents

Do not wear wool, silk, leather or latex gloves and cover exposed skin

Perform blank extractions alongside sample extractions

Separate ancient and modern laboratories and prevent sharing of reagents

Mass Spectrometry

Perform injection blanks prior to and during LC-MS/MS runs

Perform experimental replications or multiple MS/MS runs for validating results

Peptide and Protein Identification

Include soil, microbial and/or common contaminants in *in silico* databases

Include diagenetic protein modifications in search strategies

Support amino acid sequences with more than one MS/MS PSM

Present statistical parameters or experiments for error-tolerant or *de novo* strategies

Data Interpretation and Authentication

Perform additional biomolecular techniques to support findings, where applicable

Report the presence of diagenetic PTMs and non-tryptic cleavages

Data Integration and Sharing

Share raw and processed mass spectrometric data on public repositories

Summary Box 1. Crucial aspects of a palaeoproteomics workflow, from sample selection to data sharing.

Supplementary Table 1. Reporting of extraction blanks, injection blanks, evidence of protein degradation and MS data reporting in MS/MS-based anciprotein analysis publications. Extraction and injection blanks are marked as present when they are explicitly mentioned in the manuscript; if marked as absent, this does not necessarily suggest that these blanks were in fact not run or analyzed in the experiment, but they are not reported. MALDI-TOF-MS, MALDI-TOF/TOF-MS and antibody-based studies are not included. Accession numbers in the final column refer to datasets stored in ProteomeXchange, otherwise the name of other repositories is given; in one case this refers to a university-based ftp page that can be accessed using details provided in the relevant paper. [1]Degradation noted by the presence of smeared gels.

Supplementary Table 2. Demonstration of misleading species assignments in Mascot outputs. Of the top 20 eukaryotic proteins (ranked by score) identified from sheep tooth cementum, only 4 are assigned to sheep. Although the protein identifications themselves are expected for bone/dentine/cementum, misleading species assignments to *Bos taurus, Homo sapiens* and *Mus musculus* are made when the SwissProt database lacks the relevant sheep reference protein.

References

1. Cappellini, E. *et al.* Proteomic analysis of a pleistocene mammoth femur reveals more than one hundred ancient bone proteins. *J. Proteome Res.* **11,** 917–926 (2012).
2. Cappellini, E. *et al.* Resolution of the type material of the Asian elephant, Elephas maximus Linnaeus, 1758 (Proboscidea, Elephantidae). *Zool. J. Linn. Soc.* **170,** 222–232 (2014).
3. Warinner, C. *et al.* Pathogens and host immunity in the ancient human oral cavity. *Nat. Genet.* **46,** 336–344 (2014).
4. Welker, F. *et al.* Ancient proteins resolve the evolutionary history of Darwin/'s South American ungulates. *Nature* **522,** 81–84 (2015).
5. Rybczynski, N. *et al.* Mid-Pliocene warm-period deposits in the High Arctic yield insight into camel evolution. *Nat. Commun.* **4,** 1550 (2013).
6. Cleland, T. P., Schroeter, E. R., Feranec, R. S. & Vashishth, D. Peptide sequences from the first Castoroides ohioensis skull and the utility of old museum collections for palaeoproteomics. *Proc. Biol. Sci.* **283,** (2016).
7. Welker, F. *et al.* Palaeoproteomic evidence identifies archaic hominins associated with the Châtelperronian at the Grotte du Renne. *Proc. Natl. Acad. Sci. U. S. A.* **113,** 11162–11167 (2016).
8. Demarchi, B. *et al.* Protein sequences bound to mineral surfaces persist into deep time. *Elife* **5,** e17092 (2016).
9. Hill, R. C. *et al.* Preserved Proteins from Extinct Bison latifrons Identified by Tandem Mass Spectrometry; Hydroxylysine Glycosides are a Common Feature of Ancient Collagen. *Mol. Cell. Proteomics* **14,** 1946–1958 (2015).
10. Cleland, T. P., Schroeter, E. R. & Schweitzer, M. H. Biologically and diagenetically derived peptide modifications in moa collagens. *Proceedings of the Royal Society B* **282,** 20150015 (2015).
11. Mikšík, I., Sedláková, P., Pataridis, S., Bortolotti, F. & Gottardo, R. Proteins and their modifications in a medieval mummy. *Protein Sci.* (2016). doi:10.1002/pro.3024
12. Warinner, C. *et al.* Direct evidence of milk consumption from ancient human dental calculus. *Sci. Rep.* **4,** 7104 (2014).
13. Corthals, A. *et al.* Detecting the immune system response of a 500 year-old Inca mummy. *PLoS One* **7,** e41244 (2012).
14. Maixner, F. *et al.* Paleoproteomic study of the Iceman's brain tissue. *Cell. Mol. Life Sci.* **70,** 3709–3722 (2013).
15. Kendall, R., Hendy, J., Collins, M. J., Millard, A. R. & Gowland, R. L. Poor preservation of antibodies in archaeological human bone and dentine. *STAR: Science & Technology of Archaeological Research* **2,** 15–24 (2016).
16. Hendy, J. *et al.* The challenge of identifying tuberculosis proteins in archaeological tissues. *J. Archaeol. Sci.* **66,** 146–153 (2016).
17. Buckley, M., Collins, M. J., 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.* **23,** 3843–3854 (2009).
18. von Holstein, I. C. C. *et al.* Searching for Scandinavians in pre-Viking Scotland: molecular fingerprinting of Early Medieval combs. *J. Archaeol. Sci.* **41,** 1–6 (2014).
19. Stewart, N. A. *et al.* The identification of peptides by nanoLC-MS/MS from human surface tooth

enamel following a simple acid etch extraction. *RSC Adv.* **6,** 61673–61679 (2016).
20. Stewart, J. R. M., Allen, R. B., Jones, A. K. G., Penkman, K. E. H. & Collins, M. J. ZooMS: making eggshell visible in the archaeological record. *J. Archaeol. Sci.* **40,** 1797–1804 (2013).
21. Warinner, C., Speller, C. & Collins, M. J. A new era in palaeomicrobiology: prospects for ancient dental calculus as a long-term record of the human oral microbiome. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **370,** 20130376 (2015).
22. Cappellini, E. *et al.* A multidisciplinary study of archaeological grape seeds. *Naturwissenschaften* **97,** 205–217 (2010).
23. Shevchenko, A. *et al.* Proteomics identifies the composition and manufacturing recipe of the 2500-year old sourdough bread from Subeixi cemetery in China. *J. Proteomics* **105,** 363–371 (2014).
24. Yang, Y. *et al.* Proteomics evidence for kefir dairy in Early Bronze Age China. *J. Archaeol. Sci.* **45,** 178–186 (2014).
25. Xie, M. *et al.* Identification of a dairy product in the grass woven basket from Gumugou Cemetery (3800 BP, northwestern China). *Quat. Int.* **426,** 158–165 (2016).
26. Craig, O. E. *et al.* Detecting milk proteins in ancient pots. *Nature* **408,** 312 (2000).
27. Solazzo, C., Fitzhugh, W. W., Rolando, C. & Tokarski, C. Identification of protein remains in archaeological potsherds by proteomics. *Anal. Chem.* **80,** 4590–4597 (2008).
28. Buckley, M., Melton, N. D. & Montgomery, J. Proteomics analysis of ancient food vessel stitching reveals> 4000-year-old milk protein. *Rapid Commun. Mass Spectrom.* **27,** 531–538 (2013).
29. Dallongeville, S. *et al.* Proteomics applied to the authentication of fish glue: application to a 17 th century artwork sample. *Analyst* **138,** 5357–5364 (2013).
30. Dallongeville, S. *et al.* Identification of animal glue species in artworks using proteomics: application to a 18th century gilt sample. *Anal. Chem.* **83,** 9431–9437 (2011).
31. Solazzo, C. *et al.* Identification of the earliest collagen- and plant-based coatings from Neolithic artefacts (Nahal Hemar cave, Israel). *Sci. Rep.* **6,** 31053 (2016).
32. Hynek, R., Kuckova, S. & Hradilova, J. Matrix‐assisted laser desorption/ionization time‐of‐flight mass spectrometry as a tool for fast identification of protein binders in color layers of paintings. *Rapid Communications in mass spectrometry* (2004). doi:10.1002/rcm.1
33. Tokarski, C., Martin, E., Rolando, C. & Cren-Olivé, C. Identification of proteins in renaissance paintings by proteomics. *Anal. Chem.* **78,** 1494–1502 (2006).
34. Tripković, T. *et al.* Electrospray ionization linear trap quadrupole Orbitrap in analysis of old tempera paintings: application to nineteenth-century Orthodox icons. *Eur. J. Mass Spectrom.* **21,** 679–692 (2015).
35. Brandt, L. Ø. *et al.* Species identification of archaeological skin objects from Danish bogs: comparison between mass spectrometry-based peptide sequencing and microscopy-based methods. *PLoS One* **9,** e106875 (2014).
36. Gong, Y., Li, L., Gong, D., Yin, H. & Zhang, J. Biomolecular Evidence of Silk from 8,500 Years Ago. *PLoS One* **11,** e0168042 (2016).
37. Fiddyment, S. *et al.* Animal origin of 13th-century uterine vellum revealed using noninvasive peptide fingerprinting. *Proc. Natl. Acad. Sci. U. S. A.* **112,** 15066–15071 (2015).
38. Kuckova, S., Hynek, R. & Kodicek, M. Application of peptide mass mapping on proteins in historical mortars. *J. Cult. Herit.* **10,** 244–247 (2009).
39. Krizkova, M. C., Kuckova, S. H., Santrucek, J. & Hynek, R. Peptide mass mapping as an effective

tool for historical mortar analysis. *Construction and Building Materials* **50,** 219–225 (2014).
40. Rao, H., Li, B., Yang, Y., Ma, Q. & Wang, C. Proteomic identification of organic additives in the mortars of ancient Chinese wooden buildings. *Anal. Methods* (2014). doi:10.1039/C4AY01766H
41. Oonk, S., Cappellini, E. & Collins, M. J. Soil proteomics: An assessment of its potential for archaeological site interpretation. *Org. Geochem.* **50,** 57–67 (2012).
42. Abelson, P. H. Amino acids in fossils. *Science* **119,** 576 (1954).
43. Hare, E. & Abelson, P. H. Racemization of amino acids in fossil shells. *Year B. Carnegie Inst. Wash.* **66,** 526–528 (1968).
44. Schroeder, R. A. & Bada, J. L. A review of the geochemical applications of the amino acid racemization reaction. *Earth-Sci. Rev.* **12,** 347–391 (1976).
45. Hoering, T. C. The organic constituents of fossil mollusk shells. in *Biogeochem. Amino Acids Conference papers* 193–201 (bcin.ca, 1978).
46. *Perspectives in Amino Acid and Protein Geochemistry*. (Oxford University Press Inc, 2001).
47. Miller, M. F., 2nd & Wyckoff, R. W. Proteins in dinosaur bones. *Proc. Natl. Acad. Sci. U. S. A.* **60,** 176–178 (1968).
48. Towe, K. M. Preserved organic ultrastructure: an unreliable indicator for Paleozoic amino acid biogeochemistry. *Biogeochemistry of amino acids* 558 (1980).
49. Fu, Q. *et al.* A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr. Biol.* **23,** 553–559 (2013).
50. Renaud, G., Slon, V., Duggan, A. T. & Kelso, J. Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. *Genome Biol.* **16,** 224 (2015).
51. Buckley, M., Warwood, S., van Dongen, B., Kitchener, A. C. & Manning, P. L. A fossil protein chimera; difficulties in discriminating dinosaur peptide sequences from modern cross-contamination. *Proc. Biol. Sci.* **284,** (2017).
52. American Society for Biochemistry and Molecular Biology. Instructions to Authors Regarding Required Manuscript Content and Publication Guidelines for Molecular and Cellular Proteomics. (2017). Available at: http://www.mcponline.org/site/misc/ms_guidelines.xhtml. (Accessed: 15th June 2017)
53. Gilbert, M. T. P., Bandelt, H.-J., Hofreiter, M. & Barnes, I. Assessing ancient DNA studies. *Trends Ecol. Evol.* **20,** 541–544 (2005).
54. van Doorn, N. L., Hollund, H. & Collins, M. J. A novel and non-destructive approach for ZooMS analysis: ammonium bicarbonate buffer extraction. *Archaeol. Anthropol. Sci.* **3,** 281–289 (2011).
55. Manfredi, M. *et al.* Method for Noninvasive Analysis of Proteins and Small Molecules from Ancient Objects. *Anal. Chem.* **89,** 3310–3317 (2017).
56. Teasdale, M. D. *et al.* The York Gospels: a one thousand year biological palimpsest. *bioRxiv* 146324 (2017). doi:10.1101/146324
57. Tokarski, C., Martin, E., Rolando, C. & Cren-Olivé, C. Identification of proteins in renaissance paintings by proteomics. *Anal. Chem.* **78,** 1494–1502 (2006).
58. Hedges - Archaeometry, R. & 2002. Bone diagenesis: an overview of processes. *Wiley Online Library* (2002).
59. Collins, M. J. *et al.* The survival of organic matter in bone: a review. *Archaeometry* **44,** 383–394 (2002).
60. Demarchi, B. *et al.* Intra-crystalline protein diagenesis (IcPD) in Patella vulgata. Part I: Isolation and

testing of the closed system. *Quat. Geochronol.* **16,** 144–157 (2013).
61. Mackie, M. *et al.* Preservation of the metaproteome: variability of protein preservation in ancient dental calculus. *STAR: Science & Technology of Archaeological Research* **3,** 74–86 (2017).
62. Stankiewicz, A. B. *et al.* Recognition of Chitin and Proteins in Invertebrate Cuticles Using Analytical Pyrolysis/Gas Chromatography and Pyrolysis/Gas Chromatography/Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **10,** 1747–1757 (1996).
63. Saitta, E. T. *et al.* Low fossilization potential of keratin protein revealed by experimental taphonomy. *Palaeontology* **60,** 547–556 (2017).
64. Keck, S. & Peters, T. Identification of Protein-Containing Paint Media by Quantitative Amino Acid Analysis. *Stud. Conserv.* **14,** 75–82 (1969).
65. Wyckoff, R. W., Wagner, E., Matter, P., 3rd & Doberenz, A. R. COLLAGEN IN FOSSIL BONE. *Proc. Natl. Acad. Sci. U. S. A.* **50,** 215–218 (1963).
66. Brock, F., Geoghegan, V., Thomas, B., Jurkschat, K. & Higham, T. F. G. Analysis of Bone 'Collagen' Extraction Products for Radiocarbon Dating. *Radiocarbon* **55,** 445–463 (2013).
67. van Doorn, N. L., Wilson, J., Hollund, H., Soressi, M. & Collins, M. J. Site-specific deamidation of glutamine: a new marker of bone collagen deterioration. *Rapid Commun. Mass Spectrom.* **26,** 2319–2327 (2012).
68. Simpson, J. P. *et al.* The effects of demineralisation and sampling point variability on the measurement of glutamine deamidation in type I collagen extracted from bone. *J. Archaeol. Sci.* **69,** 29–38 (2016).
69. Schellmann, N. C. Animal glues: a review of their key properties relevant to conservation. *Stud. Conserv.* **52,** 55–66 (2007).
70. Hodge, K., Have, S. T., Hutton, L. & Lamond, A. I. Cleaning up the masses: exclusion lists to reduce contamination with HPLC-MS/MS. *J. Proteomics* **88,** 92–103 (2013).
71. Wiktorowicz, C. J., Arnold, B., Wiktorowicz, J. E., Murray, M. L. & Kurosky, A. Hemorrhagic fever virus, human blood, and tissues in Iron Age mortuary vessels. *J. Archaeol. Sci.* **78,** 29–39 (2017).
72. Bergeron, É. *et al.* Recovery of Recombinant Crimean Congo Hemorrhagic Fever Virus Reveals a Function for Non-structural Glycoproteins Cleavage by Furin. *PLoS Pathog.* **11,** e1004879 (2015).
73. Welker, F. *et al.* Variations in glutamine deamidation for a Châtelperronian bone assemblage as measured by peptide mass fingerprinting of collagen. *STAR: Science & Technology of Archaeological Research* **3,** 15–27 (2017).
74. Sawafuji, R. *et al.* Proteomic profiling of archaeological human bone. *Royal Society Open Science* **4,** 161004 (2017).
75. Sykes, G. A., Collins, M. J. & Walton, D. I. The significance of a geochemically isolated intracrystalline organic fraction within biominerals. *Org. Geochem.* **23,** 1059–1065 (1995).
76. Penkman, K. E. H., Kaufman, D. S., Maddy, D. & Collins, M. J. Closed-system behaviour of the intra-crystalline fraction of amino acids in mollusc shells. *Quat. Geochronol.* **3,** 2–25 (2008).
77. Damgaard, P. B. *et al.* Improving access to endogenous DNA in ancient bones and teeth. *Sci. Rep.* **5,** 11184 (2015).
78. Korlević, P. *et al.* Reducing microbial and human contamination in DNA extractions from ancient bones and teeth. *Biotechniques* **59,** 87–93 (2015).
79. Ginolhac, A. *et al.* Improving the performance of true single molecule sequencing for ancient DNA. *BMC Genomics* **13,** 177 (2012).

80. Demarchi, B. & Collins, M. Amino Acid Racemization Dating. in *Encyclopedia of Scientific Dating Methods* 1–22 (Springer Netherlands, 2014).
81. Jiang, X. *et al.* Method development of efficient protein extraction in bone tissue for proteome analysis. *J. Proteome Res.* **6,** 2287–2294 (2007).
82. Wadsworth, C. & Buckley, M. Proteome degradation in fossils: investigating the longevity of protein survival in ancient bone. *Rapid Commun. Mass Spectrom.* **28,** 605–615 (2014).
83. Wadsworth, C. & Buckley, M. Characterization of Proteomes Extracted through Collagen-based Stable Isotope and Radiocarbon Dating Methods. *J. Proteome Res.* **17,** 429–439 (2018).
84. Shevchenko, A., Tomas, H., Havlis, J., Olsen, J. V. & Mann, M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protoc.* **1,** 2856–2860 (2006).
85. Wiśniewski, J. R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nat. Methods* **6,** 359–362 (2009).
86. Warinner, C. *et al.* Pathogens and host immunity in the ancient human oral cavity. *Nat. Genet.* **46,** 336–344 (2014).
87. Fischer, R. & Kessler, B. M. Gel-aided sample preparation (GASP)-A simplified method for gel-assisted proteomic sample generation from protein extracts and intact cells. *Proteomics* (2014). doi:10.1002/pmic.201400436
88. Colonese, A. C. *et al.* New criteria for the molecular identification of cereal grains associated with archaeological artefacts. *Sci. Rep.* **7,** 6633 (2017).
89. Procopio, N. & Buckley, M. Minimizing Laboratory-Induced Decay in Bone Proteomics. *J. Proteome Res.* **0,** null (2016).
90. Schroeter, E. R., DeHart, C. J., Schweitzer, M. H., Thomas, P. M. & Kelleher, N. L. Bone protein 'extractomics': comparing the efficiency of bone protein extractions of Gallus gallus in tandem mass spectrometry, with an eye towards paleoproteomics. *PeerJ* **4,** e2603 (2016).
91. Taylor, C. F. *et al.* Guidelines for reporting the use of mass spectrometry in proteomics. *Nat. Biotechnol.* **26,** 860–861 (2008).
92. Taylor, G. K. & Goodlett, D. R. Rules governing protein identification by mass spectrometry. *Rapid Commun. Mass Spectrom.* **19,** 3420 (2005).
93. Picotti, P., Aebersold, R. & Domon, B. The implications of proteolytic background for shotgun proteomics. *Mol. Cell. Proteomics* **6,** 1589–1598 (2007).
94. Burkhart, J. M., Schumbrutzki, C., Wortelkamp, S., Sickmann, A. & Zahedi, R. P. Systematic and quantitative comparison of digest efficiency and specificity reveals the impact of trypsin quality on MS-based proteomics. *J. Proteomics* **75,** 1454–1462 (2012).
95. Kim, J.-S., Monroe, M. E., Camp, D. G., 2nd, Smith, R. D. & Qian, W.-J. In-source fragmentation and the sources of partially tryptic peptides in shotgun proteomics. *J. Proteome Res.* **12,** 910–916 (2013).
96. The Global Proteome Machine Organization. cRAP protein sequences. (2017). Available at: http://www.thegpm.org/crap/. (Accessed: 1st June 2017)
97. Mellacheruvu, D. *et al.* The CRAPome: a contaminant repository for affinity purification-mass spectrometry data. *Nat. Methods* **10,** 730–736 (2013).
98. Penkman, K. & Kaufman, D. Amino acid geochronology: Recent perspectives. *Quat. Geochronol.* 1–2 (2013).
99. Welker, F. *et al.* Middle Pleistocene protein sequences from the rhinoceros genus Stephanorhinus

and the phylogeny of extant and extinct Middle/Late Pleistocene Rhinocerotidae. *PeerJ* **5,** e3033 (2017).
100. Breci, L. A., Tabb, D. L., Yates, J. R., 3rd & Wysocki, V. H. Cleavage N-terminal to proline: analysis of a database of peptide tandem mass spectra. *Anal. Chem.* **75,** 1963–1971 (2003).
101. Schroeter, E. R. *et al.* Expansion for the Brachylophosaurus canadensis Collagen I Sequence and Additional Evidence of the Preservation of Cretaceous Protein. *J. Proteome Res.* **16,** 920–932 (2017).
102. Scientific Data. Recommended Data Repositories. (2017). Available at: https://www.nature.com/sdata/policies/repositories. (Accessed: 30th May 2017)
103. Vizcaíno, J. A. *et al.* ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat. Biotechnol.* **32,** 223–226 (2014).
104. Anagnostou, P. *et al.* When data sharing gets close to 100%: what human paleogenetics can teach the open science movement. *PLoS One* **10,** e0121409 (2015).
105. Welker, F. The Palaeoproteomic Identification of Pleistocene Hominin Skeletal Remains: Towards a Biological Understanding of the Middle to Upper Palaeolithic Transition. (Max-Planck-Institute for Evolutionary Anthropology, 2017).
104. Asara, J. M., Schweitzer, M. H., Freimark, L. M., Phillips, M. & Cantley, L. C. Protein sequences from mastodon and Tyrannosaurus rex revealed by mass spectrometry. *Science* **316,** 280–285 (2007).
105. Schweitzer, Mary H., Wenxia Zheng, Chris L. Organ, Recep Avci, Zhiyong Suo, Lisa M. Freimark, Valerie S. Lebleu, et al. 2009. "Biomolecular Characterization and Protein Sequences of the Campanian Hadrosaur B. Canadensis." *Science* 324 (5927):626–31.
106. Hollemeyer, K., Altmeyer, W., Heinzle, E. & Pitra, C. Species identification of Oetzi's clothing with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry based on peptide pattern similarities of hair digests. *Rapid Commun. Mass Spectrom.* **22,** 2751–2767 (2008).
107. Leo, G. *et al.* Proteomic strategies for the identification of proteinaceous binders in paintings. *Anal. Bioanal. Chem.* **395,** 2269–2280 (2009).
108. Leo, G. *et al.* Deamidation at asparagine and glutamine as a major modification upon deterioration/aging of proteinaceous binders in mural paintings. *Anal. Chem.* **83,** 2056–2064 (2011).
109. Buckley, M., Larkin, N. & Collins, M. Mammoth and Mastodon collagen sequences; survival and utility. *Geochim. Cosmochim. Acta* **75,** 2007–2016 (2011).
110. Porto, I. M. *et al.* Recovery and identification of mature enamel proteins in ancient teeth. *Eur. J. Oral Sci.* **119 Suppl 1,** 83–87 (2011).
111. Boros-Major, A. *et al.* New perspectives in biomolecular paleopathology of ancient tuberculosis: a proteomic approach. *J. Archaeol. Sci.* **38,** 197–201 (2011).
112. Hong, C. *et al.* Identification of milk component in ancient food residue by proteomics. *PLoS One* **7,** e37053 (2012).
113. Toniolo, L., D'Amato, A., Saccenti, R., Gulotta, D. & Righetti, P. G. The Silk Road, Marco Polo, a Bible and its proteome: a detective story. *J. Proteomics* **75,** 3365–3373 (2012).
114. Orlando, L. *et al.* Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature* **499,** 74–78 (2013).
115. Buckley, M., Melton, N. D. & Montgomery, J. Proteomics analysis of ancient food vessel stitching reveals> 4000-year-old milk protein. *Rapid Commun. Mass Spectrom.* **27,** 531–538 (2013).
116. Mikšík, I. *et al.* Prince Cangrande's Collagen: Study of Protein Modification on the Mummy of

the Lord of Verona, Italy (1291–1329 AD). *Chromatographia* **77,** 1503–1510 (2014).
117. Bona, A. *et al.* Mass spectrometric identification of ancient proteins as potential molecular biomarkers for a 2000-year-old osteogenic sarcoma. *PLoS One* **9,** e87215 (2014).
118. Wadsworth, C. & Buckley, M. Proteome degradation in fossils: investigating the longevity of protein survival in ancient bone. *Rapid Commun. Mass Spectrom.* **28,** 605–615 (2014).
119. Buckley, M. & Wadsworth, C. Proteome degradation in ancient bone: Diagenesis and phylogenetic potential. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **416,** 69–79 (2014).
120. Rao, H. *et al.* Proteomic identification of adhesive on a bone sculpture-inlaid wooden artifact from the Xiaohe Cemetery, Xinjiang, China. *J. Archaeol. Sci.* **53,** 148–155 (2015).
121. Bleicher, N., Kelstrup, C., Olsen, J. V. & Cappellini, E. Molecular evidence of use of hide glue in 4th millennium BC Europe. *J. Archaeol. Sci.* **63,** 65–71 (2015).
122. Villa, P. *et al.* A Milk and Ochre Paint Mixture Used 49,000 Years Ago at Sibudu, South Africa. *PLoS One* **10,** e0131273 (2015).
123. Buckley, M. Ancient collagen reveals evolutionary history of the endemic South American 'ungulates'. *Proceedings of the Royal Society of London B: Biological Sciences* **282,** (2015).
124. Buckley, M. *et al.* Collagen Sequence Analysis of the Extinct Giant Ground Sloths Lestodon and Megatherium. *PLoS One* **10,** e0144793 (2015).
125. Barker, A. *et al.* Validation of a Non-Targeted LC-MS Approach for Identifying Ancient Proteins: Method Development on Bone to Improve Artifact Residue Analysis. *Ethnobiology Letters* **6,** 162–174 (2015).
126. Brown, S. *et al.* Identification of a new hominin bone from Denisova Cave, Siberia using collagen fingerprinting and mitochondrial DNA analysis. *Sci. Rep.* **6,** 23559 (2016).
127. Jones, J. *et al.* Identification of proteins from 4200-year-old skin and muscle tissue biopsies from ancient Egyptian mummies of the first intermediate period shows evidence of acute inflammation and severe immune response. *Philos. Trans. A Math. Phys. Eng. Sci.* **374,** (2016).
128. Mai, H., Yang, Y., Jiang, H., Wang, B. & Wang, C. Investigating the materials and manufacture of Jinzi: The lining of Futou (Chinese traditional male headwear) from the Astana Cemeteries, Xinjiang, China. *J. Cult. Herit.* **27,** 116–124 (2017).
129. Shevchenko, A. *et al.* Open sesame: Identification of sesame oil and oil soot ink in organic deposits of Tang Dynasty lamps from Astana necropolis in China. *PLoS One* **12,** e0158636 (2017).
130. Mai, H., Yang, Y., Jiang, H., Wang, B. & Wang, C. Investigating the materials and manufacture of Jinzi: The lining of Futou (Chinese traditional male headwear) from the Astana Cemeteries, Xinjiang, China. *J. Cult. Herit.* **27,** 116–124 (2017).
131. Wadsworth, C. *et al.* Comparing ancient DNA survival and proteome content in 69 archaeological cattle tooth and bone samples from multiple European sites. *J. Proteomics* **158,** 1–8 (2017).
132. Fresnais, M., Richardin, P. & Sepúlveda, M. Omics for Precious Rare Biosamples: Characterization of Ancient Human Hair by a Proteomic Approach. *OMICS* (2017).
133. Stewart, N. A., Gerlach, R. F., Gowland, R. L., Gron, K. J. & Montgomery, J. Sex determination of human remains from peptides in tooth enamel. *Proc. Natl. Acad. Sci. U. S. A.* **114,** 13649–13654 (2017).
134. Barbieri, R. *et al.* Paleoproteomics of the Dental Pulp: The plague paradigm. *PLoS One* **12,** e0180552 (2017).

135. Yu, X. *et al.* Mass spectrometry analysis of textile used in decorating the coronet of Empress Xiao of the Sui Dynasty (581–618 A.D.). *J. Cult. Herit.* **25,** 185–188 (2017).
136. Zilberstein, G., Maor, U., Baskin, E., D'Amato, A. & Righetti, P. G. Unearthing Bulgakov's trace proteome from the Master i Margarita manuscript. *J. Proteomics* **152,** 102–108 (2017).
137. Wadsworth, C. & Buckley, M. Characterization of Proteomes Extracted through Collagen-based Stable Isotope and Radiocarbon Dating Methods. *J. Proteome Res.* **17,** 429–439 (2018).
138. Procopio, N., Chamberlain, A. T. & Buckley, M. Exploring biological and geological age-related changes through variations in intra- and inter-tooth proteomes of ancient dentine. *J. Proteome Res.* (2018). doi:10.1021/acs.jproteome.7b00648
139. D'Amato, A., Zilberstein, G., Zilberstein, S., Compagnoni, B. L. & Righetti, P. G. Of mice and men: Traces of life in the death registries of the 1630 plague in Milano. *J. Proteomics* (2018). doi:10.1016/j.jprot.2017.11.028
140. Krizova, I. *et al.* Comparison of analytical tools appropriate for identification of proteinaceous additives in historical mortars. *Anal. Bioanal. Chem.* **410,** 189–200 (2018).