CHEMICAL REVIEWS

pubs.acs.org/CR

Paleoproteomics

Christina Warinner,* Kristine Korzow Richter, and Matthew J. Collins
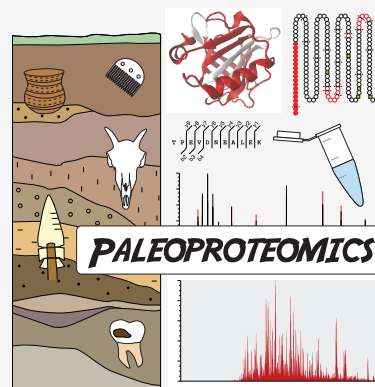
Read Online

ACCESS | 📊 Metrics & More | 📖 Article Recommendations

ABSTRACT: Paleoproteomics, the study of ancient proteins, is a rapidly growing field at the intersection of molecular biology, paleontology, archaeology, paleoecology, and history. Paleoproteomics research leverages the longevity and diversity of proteins to explore fundamental questions about the past. While its origins predate the characterization of DNA, it was only with the advent of soft ionization mass spectrometry that the study of ancient proteins became truly feasible. Technological gains over the past 20 years have allowed increasing opportunities to better understand preservation, degradation, and recovery of the rich bioarchive of ancient proteins found in the archaeological and paleontological records. Growing from a handful of studies in the 1990s on individual highly abundant ancient proteins, paleoproteomics today is an expanding field with diverse applications ranging from the taxonomic identification of highly fragmented bones and shells and the phylogenetic resolution of extinct species to the exploration of past cuisines from dental calculus and pottery food crusts and the characterization of past diseases. More broadly, these studies have opened new doors in understanding past human−animal interactions, the reconstruction of past environments and environmental changes, the expansion of the hominin fossil record through large scale screening of nondiagnostic bone fragments, and the phylogenetic resolution of the vertebrate fossil record. Even with these advances, much of the ancient proteomic record still remains unexplored. Here we provide an overview of the history of the field, a summary of the major methods and applications currently in use, and a critical evaluation of current challenges. We conclude by looking to the future, for which innovative solutions and emerging technology will play an important role in enabling us to access the still unexplored "dark" proteome, allowing for a fuller understanding of the role ancient proteins can play in the interpretation of the past.

CONTENTS

1. INTRODUCTION

The study of ancient proteins is at once a very old and a very young field. First explored in the 1930s[1] and later formulated as "paleobiochemistry" in the 1950s,[2] the early history of ancient protein research is deeply rooted in the fields of chemistry, anthropology, and geology. However, it was only following the application of soft ionization mass spectrometry in the early 2000s[3] that the study of ancient protein sequences became truly feasible, developing into the field known as paleoproteomics today.

Ancient protein research is now advancing at a rapid pace, and its application includes the study of a wide range of archaeological, historical, and paleontological remains and materials.[4−7] Often compared to its sister field of paleogenomics, paleoproteomics is not yet as developed in scale or scope, but its demonstrated success in retrieving biomolecular sequence data from samples beyond the limit of ancient DNA (aDNA) and its ability to characterize specific tissues and biological processes make it particularly valuable and give it enhanced interpretive nuance.

To probe past life using biomolecules, in the time scale of millions of years, proteins are likely to be our best resource. Proteins are found in almost all biological tissues, and before the age of plastics they also made up a large proportion of the material culture produced by human societies around the world. Proteins persist long beyond their biological function, becoming foods, textiles, building materials, paints, and glues. The remnants of these past materials and activities have become incorporated into the historical and archaeological records, just as the remains of humans, animals, and plants have become integrated into the bio- and geosphere, where they can remain accessible into deep time.

Although proteins decay, nitrogen recycling is not completely efficient, and in protected environments (e.g.,

bones, teeth, eggshell) proteins can persist for millions of years or more. Protein fragments are recognizable in fossils (e.g., seeds, bone), worked biological remains, (e.g., wood, textiles, archaeological and art historical artifacts), as residues on cooking vessels, and also entrapped within soils and sediments. There is more protein nitrogen in this "dead pool" than there is in all the living cells on earth.[8] Encoded by DNA, proteins pack the same amount of sequence information into approximately one-sixth the number of atoms. For example, a 50 bp fragment of DNA (30.4 kDa) has a larger mass than many intact proteins, including β-lactoglobulin (18.4 kDa), hemoglobin (15.9 kDa), and amelogenin (24.1 kDa). Protein folding and aggregation further protect proteins from chemical attack and facilitate entrapment. With fewer atoms, fewer chemical bonds, and a more compact structure, proteins consequently fall apart more slowly than DNA. However, the greater range of reactive species and our limited ability to recover direct information about their state of decay mean that ancient proteins stretch the limits of our understanding of decay processes and diagenetic modification. Yet the results are hardly esoteric, as modifications associated with ancient proteins have relevance for understanding aging and diseased tissues, and are induced during the production and consumption of protein-containing materials and foods.

In this review, we discuss the history of paleoproteomics, the revolutionary change brought about by mass spectrometry, and the methods and applications currently in use. We further detail the main challenges facing ancient protein research today and offer perspectives on future directions in the field.

1.1. Proteins as a Bioarchive of the Past

Proteins are long-lived biomolecules capable of surviving over millions of years.[9,10] They routinely outlast even the oldest surviving DNA,[11−14] and their full longevity has yet to be determined.[15−17] Although proteins do not persist into deep time as long as lipids,[18] their sequence diversity makes them more informative, and consequently proteins represent one of our most valuable bioarchives of the past.

The longevity and biological utility of proteins derive in large part from their structure. Proteins are large biomolecules built from linear sequences of amino acids folded into complex three-dimensional forms. The 20 standard amino acids, each formed around a central carbon, contain a carboxyl group and an amino group, which form the peptide bonds linking the amino acids together into proteins, and an R group, which varies between amino acids and imparts distinct chemical properties. R groups are chemically diverse, consisting of positively charged, negatively charged, polar, and nonpolar groups that can be small, large, or structurally constrained. The sequence of amino acids making up the primary protein structure is encoded by DNA, which is then transcribed to RNA and translated into proteins using trinucleotide codon sequences for each amino acid. Because proteins are derived from the genetic code, individual proteins preserve part of the heritable genetic signal of an organism, and therefore, protein sequences can be used to make taxonomic identifications and reconstruct phylogenies.[19−21]

After protein synthesis, additional post-translational modifications (PTMs) can be made to the amino acids, changing their chemical properties.[22,23] Protein splicing, autoprocessing, conjugation, and other forms of modification further expand the biochemical complexity of proteins.[24,25] This biochemical diversity makes proteins substantially more complex than other

Table 1. Comparison of Instruments and Approaches Used in Ancient Protein Studies

	Immunological assays	Edman sequencing	MALDI-TOF	MALDI-TOF/TOF	LC−MS/MS
First use on ancient proteins	1937,[1,a] 1980,[608,b] 1984[58,c]	1990[65]	2000[3]	2005[74]	2006,[89,d] 2011[91,e]
Good for complex samples?	YES	NO	To some extent	To some extent	YES
Good for samples without reliable composition?	To some extent	NO	NO	NO	YES
Can get sequence data?	NO	YES	NO	YES	YES
Can target specific proteins	YES	NO	NO	NO	To some extent
Proteins detected in one analysis	1−5	1	1−20	1−20	100+
Feasibility for ancient samples	++	+	+++++	+++++	+++++
Reproducibility	+	++	++++	+++	+++
Relative price per sample	$$-$$$$	$$$$$	$	$$	$$$$[f]
Analysis time per sample	+++	+++++	++	++++	++++
Sample types analyzed	Any sample type	Single peptides	Sample with a few dominant proteins	Sample with a few dominant proteins	Any sample type
Examples	Hemoglobin, albumin, pathogens, silk	Osteocalcin	Collagen, keratins, silk, shell	Collagen, keratins, silk, shell	Proteomes of bone, enamel, dental calculus, artist materials

[a]Use of antisera. [b]Use of radioimmunoassay. [c]Use of ELISA. [d]Use of LC−MS/MS to identify individual ancient proteins. [e]Use of LC−MS/MS to characterize an ancient proteome of >100 proteins. [f]Depends on immunoassay design and whether antibodies are commercially available.

biomolecules, such as lipids or DNA, and it drives the folding of the linear primary amino acid chains into more complex secondary, tertiary, and quaternary structures, which form the basis of the diverse structural and functional roles of proteins.

During life, proteins are regularly degraded after their functional or structural roles are complete in order to recycle the amino acids for the creation of new proteins. While the average protein lifespan for mammals is only 1−2 days,[26] the longevity of a specific protein sequence ranges from minutes (e.g., transcription factors[27] and immune ligands[28]) to the entire lifetime of the organism (e.g., enamel[29] and eye lens proteins[30]). In addition, secreted proteins, such as hair keratins and silk proteins, form the basis of nonliving tissues and structures that can persist for centuries or more after the death of the organism.[31,32] Although proteins contain less genetic information than DNA (due to both codon redundancy and the absence of noncoding sequences), they are typically orders of magnitude more abundant, with many copies of a protein being made for every genome. Moreover, the tissue-specific expression of some proteins provides additional information about a given sample (e.g., milk vs muscle; leaf vs seed) that cannot be obtained from the genome alone.

The goal of paleoproteomics is to recover, identify, and study these proteins long after their natural lifespan, and typically after they have been extensively modified by taphonomic forces over centuries, millennia, or even millions of years as they transition from the biosphere to the lithosphere. The longevity of proteins, coupled with their biological ubiquity and diversity, makes them ideal subjects for exploring the deep and recent past, and as such they represent one of our most powerful tools for reconstructing biological and cultural history.

1.2. Origins of Paleoproteomics

The idea of using proteins to study deep time is not new. Almost 20 years before the discovery of the structure of DNA[33] and the formulation of the theoretical framework of the Central Dogma that defines the relationships between DNA, RNA, and proteins,[34,35] chemists were trying to use antisera to detect proteins in mummies and skeletons for the purpose of anthropological blood typing.[1,36,37] Interest in ancient proteins

fell away during the war years, but was revitalized in the 1950s by geophysicists working in government laboratories whose interests had shifted from bomb-making to deep time "paleobiochemistry". During the 1970s and 1980s, interest in immunological assays returned, followed by attempts to sequence ancient proteins using Edman degradation in the 1990s and eventually mass spectrometry during the 21st century (Table 1).

1.2.1. Paleobiochemistry. It is rare for any discipline to have its genesis at a single institution, yet Phil Abelson (Box 1) and the members of the Geophysical Laboratory at the Carnegie Institution of Washington, which he led from 1953 to 1971, accomplished just that. In the postwar years, they were *"free to do fundamental research unhampered by the pressures that attend work in industry and government or by the teaching load that often handicaps the university scholar"*,[38] and their time was heavily invested in pioneering the study of amino acids and proteins in deep time. The publication of Abelson's article "Paleobiochemistry: organic constituents of fossils"[2] is generally considered to mark the beginning of ancient protein studies, and in just over 1000 words he outlines a vision for what has become the field of paleoproteomics. Innovations in the detection, separation, and quantification of amino acids by members of the Geophysical Laboratory, as well as the characterization of their chirality and isotopic abundance, drove research to explore ancient proteins and their mechanisms of survival, decay, and isotopic fractionation over the next two decades. Much of this early research is detailed in the book *Biogeochemistry of Amino Acids.*[39]

One of the major outcomes of this early phase of work was the development of amino acid racemization (AAR) geo-chronology as a tool for the comparative dating of fossils,[40] a project initiated by Abelson and his protege Ed Hare (Box 1). However, this method, which examines the chiral conversion of L-amino acids into D-amino acids (either peptide-bound or free),[41] became mired in controversy following the radio-carbon redating[42] of skeletons previously analyzed by AAR that had placed the arrival of humans in North America before the last glacial maximum.[43] Although subsequent taphonomic and technical challenges further slowed the development of AAR as a relative dating technique,[44,45] recent methodological

---

Box 1. Early Pioneers in Paleoproteomics

Phil Abelson, who headed the Geophysical Laboratory at the Carnegie Institute, Washington, DC., from 1953 to 1971 was the first to identify amino acids in invertebrate and vertebrate fossils. A nuclear physicist who worked on the Manhattan Project, his pioneering work on the "paleobiochemistry" of Oligocene clams and other fossils in the 1950s[2,581−583] marks the beginning of ancient protein studies. **Ed Hare** spent almost all of his career in the Geophysical Laboratory and wrote one of the earliest Ph.D. theses on the subject of ancient proteins, focusing on the amino acid profiles of modern and fossil mussels.[584] His specialty was the chirality of fossil amino acids,[40,585−588] which led to the development of the chronometric technique of amino acid racemization dating.[589−591] **Tom Hoering** was the third member of the pioneering team based in the Geophysical Laboratory. A chemist, he specialized in mass spectroscopy and was an early pioneer of isotopic studies of organic matter[592,593] but also conducted key early experiments into the diagenesis of proteins.[594] A major impact of the trio of Abelson, Hare, and Hoering was the talent that they drew to the laboratory, who then shaped the study of amino acids and isotopes in the coming decades by scientists including Richard Mitterer, Marylin Fogel, Noreen Tuross, Mike Engle, John Wehmiller, Giff Miller, Steve Macko, Glenn Goodfriend and John Hedges.

Ralph Wyckoff was the first to seriously study amino acid decay and contamination in fossils. A skilled chemist and microscopist, Wyckoff began his career in the field of X-ray crystallography and later contributed to electron microscopy and vaccine development. He showed that while Pleistocene bones in the La Brea tar pits retained clear microscopic evidence of collagen fibers, their amino acid profiles were highly variable and differed from that of collagen,[50,51] a finding he attributed to protein degradation and amino acid instability. He found similar results for dinosaur bones[53] and mollusc shells,[52] suggesting that fossil amino acid profiles were generally unsuitable for inferring phylogeny.

Peter Wesbroek, a geologist from The Netherlands, was influenced by the work of **Margaret Jope**,[595] with whom he trained in biochemistry. After returning to The Netherlands, he established the Geobiochemistry workgroup at the University of Leiden. He studied the process of biomineralization and in doing so pioneered the use of antibodies to study ancient shell (and bone) proteins[55,596,597] as part of a wider effort to understand the role of biological systems in geological processes.[598]

Peggy Ostrom pioneered the use of mass spectrometry in ancient protein studies, innovating new methods in organic geochemistry, including stable isotope-based reconstruction of paleoecologies[599,600] and the use of isotopic analysis to identify diagenesis and contamination in fossils.[601,602] Later, she was the first to make major gains in determining ancient protein sequences by applying a variety of mass spectrometry techniques, including peptide mass fingerprinting,[3] postsource decay sequencing,[70] and tandem mass spectrometry,[69] to the study of bone osteocalcin.

improvements and an improved understanding of biomineral diagenesis are leading to a renewed interest in the approach,[46,47] which shows particular promise for dating materials that are low in organic matter, are subject to large

[14]C reservoir effects, or are beyond the limit of radiocarbon dating.[46,48,49]

1.2.2. Diagenesis, Contamination, and a Return to Immunology. Despite the early successes of the Geophysical Laboratory in recovering amino acids from fossils for the study of chemical evolution, work by Ralph Wyckoff (Box 1) and others in the 1960s made it clear that fossil proteins were highly degraded and exhibited altered amino acid profiles.[50−54] As such, they were generally unsuitable for inferring phylogeny. The absence of measurable hydroxyproline in most dinosaur bones also led Wyckoff to question the origin of the fossil-derived proteins and whether they could have resulted from the recent activity of soil microorganisms.[53]

In light of the diagenetic variability of fossil amino acids, and lacking the ability to sequence proteins directly, researchers in the 1970s and 1980s returned to immunological techniques. In 1974, Elisabeth de Vrind-de Jong (nee de Jong), Peter Westbroek and colleagues used antibodies to detect the apparent survival of epitopes in 70 Ma fossil cephalopods by immunodiffusion,[55] and Jerry Lowenstein and colleagues investigated the immunological similarity of ancient mammoths, bison, and humans to modern elephants, cattle, and humans, respectively.[56] They were able to correctly infer the systematic position of the mammoth and Tasmanian wolf using radioimmunoassay (RIA) techniques.[57] In an attempt to yield more meaningful immunological results for taxonomic systematics, an enzyme-linked immunosorbent assay (ELISA) was developed for fossil shells,[58] then later for fossil bones and teeth,[59,60] but results were difficult to replicate, and by the 1990s and 2000s immunological approaches became increasingly dogged by concerns about contamination and cross-reactivity,[61−63] as well as a lack of understanding regarding which proteins were being immunologically detected (see also ref 64).

1.2.3. Protein Sequencing. The first successful recovery of an ancient protein sequence was made by Lila Huq[65] using Edman degradation sequencing of osteocalcin from the bone of a moa (*Pachyornis elephantopus*), an extinct flightless bird. This achievement was particularly remarkable, as one of the authors of this review (M.C.) will testify, because this technology is extremely ill suited to ancient proteins. Because only a single peptide can be sequenced at a time, proteins must be isolated, digested, and purified in sufficient quantity (100 pmol) prior to analysis. Moreover, the sequencing reaction will not initiate if the reactive amino terminus is modified (e.g., by pyroglutamate), and derivatization stalls in the presence of non-α-amino acids (such as isoaspartic acid, the primary product of asparagine deamidation and aspartic acid racemization). In addition, because yields fall with each successive hour-long derivation cycle, the method is slow and limited to sequences of approximately 30−50 amino acids with high accuracy.[66,67] Even today, Edman sequencing is expensive and laborious, requiring a full day or more to sequence a single peptide using automated instruments (Table 1), although massively parallel sequencing is now being developed as a tool for single molecule sequencing.[68] Since its initial demonstration, Edman sequencing has been rarely applied to ancient proteins, but was notably used to confirm the sequence of an osteocalcin peptide identified in an early application of matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) from a 42 ka horse bone.[69] In the wake of the frustration surrounding Edman degradation, it was Peggy Ostrom and colleagues[3] who

**Peptide Mass Fingerprinting
MALDI-TOF MS**

**Proteomics
MALDI-TOF/TOF & LC-MS/MS**

Year	Peptide Mass Fingerprinting (MALDI-TOF MS)	Proteomics (MALDI-TOF/TOF & LC-MS/MS)
1946	Time-of-flight (TOF) mass spectrometry developed	
1966		Tandem mass spectrometry invented
1988	Matrix-assisted laser desorption ionization (MALDI) developed	Electrospray ionization (ESI) developed for biomolecules
1993	Development of peptide mass fingerprinting (PMF)	
1994		Database searches with MS/MS data and shotgun proteomics developed
2000	First use of PMF to identify an ancient protein (osteocalcin) by MALDI-TOF (3)	Orbitrap introduced
2001		
2002	Determination of fossil bison osteocalcin sequence by PMF and post source decay (PSD) (70)	First use of LC-MS/MS to characterize collagen in fossil bone (mammoth) (64)
2003		
2004	First use of MALDI-TOF in art (485)	
2005		Determination of Neanderthal osteocalcin sequence by PMF and MALDI-TOF/TOF (74)
2006	Determination of fossil horse osteocalcin sequence by MALDI-TOF and Edman sequencing (69)	First application of LC-MS/MS in art (89)
2007		Collagen in mastodon and *Tyrannosaurs rex* bones analyzed by LC-MS/MS (189)
2008	Osteocalcin shown to have limited potential for ancient studies (75)	Protein residues in potsherds identified by FT-ICR LC-MS/MS (467)
2009	Development of Zooarchaeology by Mass Spectrometry (ZooMS) based on collagen PMF using MALDI-TOF (72)	Filter-aided sample prepration (FASP) developed (147); Analysis of ancient enamel by MALDI-TOF/TOF (347)
2010	First use of ZooMS to distinguish sheep and goat (73)	First LC-MS/MS identification of proteins in plant macroremains (429)
2011	First use of discriminant analysis to aid ZooMS identification of taxa (fish) lacking reference sequence data (536); Development of minimally invasive AmBic sampling method for bones (155)	First high resolution characterization of an ancient bone proteome (in mammoth); first application of FASP in ancient samples (91)
2012	Characterization of deamidation in ancient collagen (556)	First identification of ancient milk proteins (105); First identification of immune proteins (from mummified tissues) (418)
2013	First application of ZooMS to cultural heritage objects (254); Development of keratin marker peptides for PMF (84); Oldest extinct megafauna (camel) characterized by ZooMS, >3Mya (10); First application of PMF on eggshell (87)	First LC-MS/MS characterization of an ancient brain (423); First identification of ancient fermented foods using LC-MS/MS (95)
2014	Expansion of collagen marker peptides for mammals (cetaceans and pinnipeds) (80)	First characterization of an ancient microbiome proteome (92); Discovery of dietary (milk) proteins in ancient dental calculus (235); Single-pot solid-phase sample preparation (SP3) developed (150)
2015	Expansion of collagen marker peptides for European fauna (257); First characterization of parchment using ZooMS; development of polyvinyl chloride (PVC) eraser method for minimally-invasive sampling (156)	Gel-aided sample prepartion (GASP) developed (148); Contested protein identification of blood vessel in dinosaur bones (16); First detailed characterization of diagenetic PTMs (in moa) (128); First revisions of extinct megafauna phylogenies based on bone collagen sequences (20)
2016	Development of combined collagen sampling methdod for ZooMS, stable isotopes, and radiocarbon (31); First identification of archaic hominid bone using high throughput ZooMS screening (291)	Guanidinium sample extraction developed (153); First LC-MS/MS investigations of ancient silk (249); Modeling of eggshell protein survival in deep time (up to 3.8 Mya); first application of guanidinium extraction method in ancient samples (9)
2017	Development of eggshell protein marker peptides for PMF (86); Development of ethyl-vinyl acetate (EVA) films for minimally-invasive sampling (295)	Development of sex typing assay using enamel amelogenin protein (94); First identification of plant proteins in dental calculus; first application of GASP for ancient samples (149)
2018	Development of bioactive films for minimally-invasive "lab-on-plate" sampling and digestion (154); Development of collagen marker peptides for freshwater and marine fish (81)	First application of SP3 to ancient samples (151); Discovery of well preserved food proteins in pottery limescale (99); Research community standards established for paleoprotoemics (4)
2019	Development of non-destructive collagen sampling method from plastic storage bags (159); Development of collagen marker peptides for reptiles (83)	First use of LC-MS/MS on ancient mollusc shell (palaeoshellomics) (197); First phylogenetic reconstruction of extinct megafauna using enamel proteome (13; 14)
2020	Development of *in situ* multi-enzyme digestion method for art (169); First use of ZooMS to identify mollusc shell (88); Development of collagen marker peptides for birds (82)	Development of protein extraction protocol compatible with ancient DNA analysis (145); First application of PRM to target an ancient protein (202); Placement of *Homo antecessor* into human phylogenetic tree using enamel proteome (172)
2021	Development for corneous beta-protein marker peptides for PMF (319); Expansion of collagen marker peptides for mammals (wild bovids) (79); Standardization of collagen nomenclature for ZooMS research (178)	First use of ZooMS and LC-MS/MS to investigate paleofeces (dog) (446); First reconstruction of trade patterns from food proteins identified in dental calculus using LC-MS/MS (102)

Figure 1. Milestones in ancient protein mass spectrometry. The broadest applications of protein mass spectrometry in archaeology today are ZooMS (zooarchaeology by mass spectrometry), which applies MALDI-TOF MS peptide mass fingerprinting to collagens, keratins, and other high abundance proteins (left), and shotgun proteomics, which uses high-resolution LC−MS/MS to identify diverse, low abundance proteins in complex mixtures (right).

made the first major breakthrough in recovering ancient protein sequences by successfully applying soft-ionization mass spectrometry to ancient proteins for the first time (Figure 1).

1.2.4. Mass Spectrometry Revolution. To this day, it seems remarkable that Peggy Ostrom's breakthrough mass spectrometry work[3,70] was not deemed sufficiently significant

to be published in any of the highest impact journals, even with the support of Geoff Eglington, the "father of organic geochemistry". Her landmark studies combined gel and immunoassay work with MALDI-TOF peptide mass finger-printing (PMF) and postsource decay (PSD) sequencing to conclusively demonstrate the presence of osteocalcin in bison bone and its survival into the late Pleistocene. This combination of immunological and mass spectrometric tools for the detection of ancient proteins, also applied by Mary Schweitzer and colleagues on mammoth bones,[64] is notable not only for the prospect of triangulating evidence but as marking the major point of transition between the two methods. While mass spectrometry has subsequently become the go-to tool for ancient protein studies, the use of immunological approaches have waned. Nevertheless, given the strengths and weaknesses of each method, there remains scope for future integration, particularly with the potential of immunoaffinity chromatography to target and enrich (or deplete) specific proteins prior to sequencing by mass spectrometry.[71]

The first widespread adoption of mass spectrometry within archaeology came with the use of MALDI-TOF MS to aid in the taxonomic discrimination of animal bones, initially sheep and goat, based on collagen peptide mass fingerprints.[72,73] While earlier efforts to characterize fossil proteins using MALDI-TOF of intact purified osteocalcin[3,69,70,74] had ultimately proven impractical due to protein degradation,[75] the application of MALDI-TOF to collagenase-digested[76] and later trypsin-digested bone collagen[72] proved a major break-through. This resulted in the development of the powerful PMF technique known as zooarchaeology by mass spectrom-etry,[6,7,77] which was given the acronym ZooMS to highlight the speed of the method and its roots in both zooarchaeology and mass spectrometry.[78] The low cost of the method and its suitability for high throughput sample processing has made it particularly powerful for many applications in archaeology, ecology, and cultural heritage (for a review see[7]), and major advances have been made over the past decade to expand the number of ZooMS markers to include a wide range of both terrestrial[79] and aquatic[80] mammals, as well as fish,[81] birds,[82] and reptiles[83] (Figure 1). Similar PMF-based approaches are also in development for additional proteins beyond collagen, including keratins[84,85] and matrix proteins in eggshell[86,87] and mollusc shell.[88]

Beyond postsource decay (PSD),[70] the use of true tandem mass spectrometry (MS/MS) to achieve a more accurate determination of peptide sequences was first achieved by John Asara and colleagues on mammoth bone using LCQ quadrupole ion trap MS/MS.[64] This was followed by the characterization of Neanderthal osteocalcin by Christina Nielsen-Marsh and colleagues using MALDI-TOF/TOF[74] and of egg proteins in Renaissance paintings by Caroline Tokarski and colleagues using nanoLC/nanoESI/Qq-TOF MS/MS.[89] These early applications of MS/MS utilized a wide variety of instrument setups, ionization techniques, and detectors, but current MS/MS analyses of ancient proteins primarily rely on LC−MS/MS systems integrating UHPLC, nano-ESI, and typically an Orbitrap[90] high performance hybrid mass spectrometer. Enhancements in speed, and more significantly for ancient samples, resolution and mass accuracy, have increased the numbers of acquired spectra and improved the success of matching these spectra to peptides. Further emerging techniques for improved ion separation, such as ion

mobility, also show great promise for improving data acquisition. However, ancient proteomes typically contain many fewer proteins, and with greater levels of modification, than equivalent modern samples. Therefore, the discipline, while still in its infancy, is currently more stifled by downstream analysis than instrumental limitations.

The major strength of MS/MS is its capacity for analyzing complex protein mixtures. The use of MS2 spectra to determine the sequence of peptides from degraded whole proteomes[91] and its integration with genomics[92] represented the next steps in the maturation of the field, followed by applications in phylogenetic interpretation,[20,21,93] sex determi-nation,[94] food preparation,[95] pathology,[96] art history,[97,98] and residue analysis,[99] among others. Major milestones in the development of mass spectrometry techniques and applications for the study of ancient proteins are highlighted in Figure 1.

2. ANCIENT PROTEINS

Today, paleoproteomics is a dynamic, fast-paced, and growing field. Regardless of the analytical techniques applied, all ancient protein studies share certain challenges and must (1) consider the formation, incorporation, and degradation processes that precede the recovery of ancient proteins, (2) apply methods to extract and prepare proteins for analysis, and (3) select appropriate analytical and interpretive strategies for character-izing ancient proteins (Figure 2). During each stage, there is progressive loss of the original proteome, an increase in chemical complexity due to diagenesis, and the addition of contaminants. Choices made in instrumentation, database selection, and data processing steps can have large impacts on the reconstruction and interpretation of ancient proteomes.

2.1. Pathways of Incorporation

Understanding the manner by which a protein was formed and how it was incorporated into a given sample is the first step in ancient protein analysis. For some samples, the manner of protein incorporation is obvious, as is the case for endogenous proteins in proteinaceous tissues such as collagens in skin and bone, keratins in hair and feathers, amelogenin in tooth enamel, and matrix proteins in mollusc shells. In such cases, the proteins comprise the tissue itself and were incorporated at the time of tissue formation. Nevertheless, some processes, such as biomineralization, are extremely complex and remain incompletely understood.[100] During biomineralization, other coassociated endogenous proteins, such as blood and plasma proteins in bone, may also become incorporated into the tissue, but the extent of biological variability related to such protein incorporation is not well studied for many tissues.

In other cases, the manner of incorporation may be less direct. Dental calculus, for example, is a calcified microbial biofilm, but in addition to bacterial proteins the dental calculus proteome is also rich in human digestive enzymes (e.g., salivary α-amylase) and immune proteins (e.g., α-s1-antitrypsin, myeloperoxidase, neutrophil defensin) that originate from saliva and gingival crevicular fluid, respectively.[92] The proteins of these fluids, which continuously bathe the teeth, become incorporated into dental calculus during periodic episodes of dental plaque mineralization. Other exogenous proteins transiently present in the oral cavity can also become incorporated during these mineralization events, including dietary proteins such as milk beta-lactoglobulin[101] and seed storage proteins.[102] Similarly, proteins within cooking vessels may become adventitiously preserved within calcified crusts
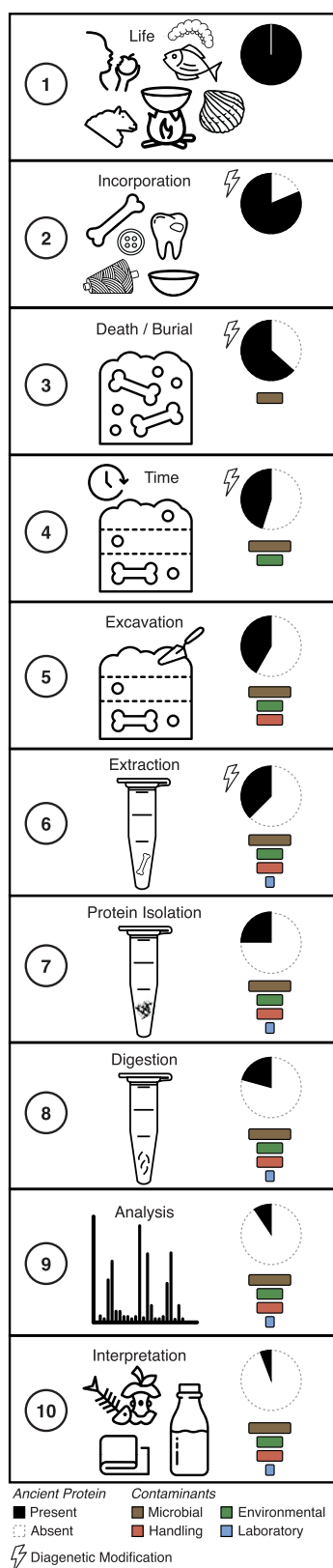
Figure 2. Conceptual stages of protein incorporation and recovery in archaeological samples. Archaeological proteins represent a small fraction of the proteins that were once present during life. Careful consideration of the full history of a sample, from incorporation to analysis, must be taken into account in order to make accurate inferences about the past.

(limescale) during evaporative comineralization[99] and during the corrosion of nearby items such as metal objects.[103,104]

In nearly all cases, mineralization is an important factor in protein long-term survival. Excluding exceptional cases from water-logged,[105,106] arid,[107,108] or very cold contexts,[109] proteins not encapsulated in a mineralized matrix generally do not persist over long periods, as has been shown experimentally for proteins applied to ceramics[110] and stone tools.[111−113] Beyond incorporation, it is also important to note that some proteomes are also altered during the incorporation process, whether by cross-linking during the tanning of leather,[114] heat denaturation during the cooking of foods,[115] or by autodegradation at the time of formation, as is the case for enamel.[116]

2.2. Processes of Decay and Diagenesis

Most organic material and proteins decay and are recycled into the environment when they are shed by a living organism or after an organism dies.[117] Degradation is primarily mediated by bacteria through enzymatic digestion, which occurs relatively quickly.[118] Experimental biodegradation of woolen fabrics and feathers has shown that even relatively robust proteins, such as keratins, which are hydrophobic and contain numerous disulfide linkages, do not survive for long under unfavorable microbial conditions.[119,120]

Only a very small percentage of the overall proteins are expected to persist in the archaeological record,[121] and those that do are generally mineralized, highly abundant, and/or have unusual properties. Type I collagen (COL1), for example, is the longest persisting bone protein, and it makes up >80% of the bone proteome (accounting for 20−30% of the mass of fresh bone), is heavily mineralized, and is arranged into a highly stable triple helix. Among dietary proteins identified in dental calculus, many are either protease inhibitors or belong to the seed storage superfamily,[102] both of which are known to be highly stable against proteolysis and thermal processing. Likewise, milk beta-lactoglobulin, which is perhaps the best attested ancient food protein, has a small molecular size and stability to changing pH levels and enzymatic degradation,[122] all properties that are known to contribute to protein survival under harsh conditions.[123] Following the initial stages of decomposition, surviving proteins are then subject to slower taphonomic processes that continue the diagenetic alterations.[124,125] In this way, nearly all ancient proteins undergo some degree of degradation or chemical damage.

Because of the diversity of proteins in terms of composition, chemical properties, size, shape, function, and incorporation (or lack thereof) into mineralized tissues, the taphonomic factors that drive post-mortem protein degradation and decay are highly variable and more poorly characterized than those for other ancient biomolecules, such as DNA. This "black box" of taphonomy is therefore an ongoing challenge for the analysis of ancient proteins, not because of a lack of research effort, but rather because of the immense complexity of the problem. Nevertheless, there are some factors that are known to play a consistent role in the protein degradation process: (1) local environment, including soil chemistry, pH, and water availability; (2) the chemical and structural composition of the matrix in which the proteins are incorporated; (3) the composition of the proteins individually and as a proteome; and (4) the local thermal history including time, temperature, and humidity.[126] These factors combine to create "diagenetiforms", or diagenetically modified protein fragments,[127]

formed through hydrolysis of peptide bonds and amino acid degradation, as well as racemization.

Protein fragmentation, the progressive, irreversible process of backbone cleavage into increasingly smaller pieces of the original protein, is among the most important forms of degradation. As the weakest covalent bond in a protein, the peptide bond is susceptible to spontaneous hydrolysis, with variable rates across the protein depending upon water accessibility to the peptide bond. This is mediated by the primary amino acid sequence, the protein secondary and tertiary structure, and surface stabilization by a mineral matrix. Protein preservation is generally better under conditions of limited water availability, such as in arid or frozen environments or where proteins are trapped in locally hydrophobic environments or in the intracrystalline fractions of biominerals.[9,126]

The other major form of protein degradation is the chemical alteration of an amino acid R group or the C or N terminus of a peptide. Such changes are myriad and incompletely characterized,[127,128] but the sheer diversity of potential chemical reactions can be appreciated by simply considering the enormous range of low molecular weight nitrogen-containing compounds formed by the diagenesis of starch storage tissues through so-called Maillard reactions.[129] Diagenesis therefore affects the chemistry—and more importantly the mass—of the affected amino acid, which can interfere with the recovery and identification of peptides using mass spectrometry. Rates of modification are again highly dependent on the primary amino acid sequence, secondary and tertiary protein structure, and surface stabilization of the protein or peptide. Indeed, it has been speculated that surface mediated preservation may promote the formation of novel condensed structures.[130] Additional modifications can be further introduced during the extraction process, either intentionally to chemically disrupt the conformation of the proteins (e.g., carbamidomethylation of cysteine by reduction and alkylation) or unintentionally through the production of undesired reactions (e.g., protein carbamylation by urea derivatives in the presence of heat).

Foundational studies on fossil invertebrates[131] and more recent studies of bone, enamel, dental calculus, and eggshell have provided insights into the range of diagenetic modifications present in ancient proteins,[9,13,92,121,132] with the most frequently identified being backbone cleavages and the deamidation of asparagine and glutamine. Other common diagenetic modifications are carboxymethylation of lysine (an advanced glycation end-product), conversion of serine to alanine, the conversion of histidine to hydroxyglutamate, the formation of N-terminus pyroglutamic acid, decomposition of arginine to ornithine, and various forms of oxidation, phosphorylation, dephosphorylation, hydroxylation, and dehydroxylation.[9,127,128] However, these represent only the forms of damage that are observable in mass spectrometry studies. Other forms of chemical modification that interfere with protein extraction and ionization are much less well understood and may mask pools of persisting but largely inaccessible proteins.[133] It is probable that many of the changes observed over time are also occurring in the kitchen, and therefore, it is worth paying attention to the expanding field of proteomics applied to food science.[134]

2.3. Methods of Recovery

In order to be detected and analyzed, proteins must first be extracted from the matrix to which they adhere or in which they are embedded. Numerous protein extraction methods are available, and their demonstrated success rates depend on the source and chemical properties of the proteins under experimental study. Compared to modern proteins, ancient protein extraction is further challenged by protein diagenetic alteration and the frequent incorporation of ancient proteins into mineral matrices. Protein loss is inevitable during this stage, both from an inability to fully "unstick" proteins from the matrix that aided their successful integration into the archaeological record and from differential recovery due to performance variation in extraction and digestion methods. Contamination may be introduced at this stage, and laboratory contaminants that have been previously observed include latex proteins from gloves, egg proteins from commercial cell lysis buffers, common laboratory reagents such as serum albumin, proteins from human sweat (e.g., dermcidin), and a wide range of keratins from human skin and sheep wool. Public lists of common laboratory contaminants, such as the common Repository of Adventitious Proteins (cRAP; https://www.thegpm.org/crap/), can aid in contaminant identification, but other potential sources of local laboratory contamination should also be considered. An awareness of potential contaminant sources and adherence to best laboratory practices is critical to mitigating laboratory contamination.[4,135]

2.3.1. Extraction Methods. Protocol development for ancient protein extraction is an active field with multiple methods in widespread use. When choosing an extraction method, sample type, size, and preservation should be considered, as well as the complexity of the proteome, the protein(s) of interest, and the amount of protein needed for analysis. The postdepositional history of the sample should also be taken into account, as well as potential chemical modifications introduced during the chosen extraction protocol.[136−141] In addition, because proteomic analyses often require less sample material than other methods, such as ancient DNA analysis, stable isotope analysis, and radiocarbon dating, protein extractions can often be performed on the leftover material or byproducts of these protocols.[139,142−145] Combining such protocols is desirable, as it reduces sampling demands on irreplaceable material.

For mineralized samples, such as enamel, bone, dental calculus, and shell, a demineralization step is generally required using either a weak acid or a chelating agent, such as ethylenediaminetetraacetic acid (EDTA). This is generally followed by protein solubilization using a variety of possible options, including heat, mechanical disruption, chaotropic agents (e.g., urea or guanidinium hydrochloride), detergents (e.g., sodium dodecyl sulfate, SDS), buffers, and salts. If the proteins are complex or are known to contain cysteines, reduction and alkylation steps are typically performed to irreversibly disrupt disulfide bonds. At this point, buffer exchange is frequently necessary to make the suspended or solubilized proteins compatible with downstream analysis, and different strategies for this are available including protocols based on the use of polyacrylamide gels,[95,146] filter-aided sample preparation (FASP),[91,102,147] gel-aided sample preparation (GASP),[148,149] single-pot solid-phase sample preparation (SP3),[150,151] or simply physical removal of the insoluble protein from the decalcification buffer in the case of collagen pseudomorphs.[152] This is then typically followed by enzymatic

digestion of the proteins into peptides, followed by peptide purification, typically using C18 resin (commercially available as StageTips [Pierce] and ZipTips [Millipore]).

For nonmineralized samples, such as artist materials (binders, glues), mummified tissues, and parchment, simplifications to the protocol can be made. Demineralization steps can be avoided, and if protein solubilization is possible in a buffer compatible with mass spectrometry (e.g., ammonium bicarbonate or guanidinium hydrochloride), buffer exchange can be avoided, which mitigates protein loss.[137,146,153] Recently, bioactive films have been developed that allow "lab-on-plate" protein extraction directly from sample material such as artwork, which further simplifies extraction of surface-available proteins.[154]

In studies focusing on a small number of highly abundant proteins of interest, such as collagens in bone and parchment, additional simplifications can be made, even for mineralized samples. Less invasive techniques can be applied to sample loosely bound proteins[155−158] and even trace proteins left behind in storage bags and containers.[144,159] However, such techniques are more susceptible to ambient contamination, and because they target unbound and largely surface proteins, the recovered proteins are likely to be more degraded.

Prior treatments or chemical exposures that interfere with protein extraction and analysis can also be mitigated in many instances. For example, synthetic adhesives, which are sometimes applied by conservators to consolidate and stabilize fragile materials, can be removed with acetone prior to protein extraction. Likewise, nonproteinaceous chemical coextractants that interfere with mass spectrometry, such as soil humic acids, can be removed from both mineralized and nonmineralized samples using sodium hydroxide (NaOH) washes during early stages of the extraction process.[160−162]

Protein contamination can be a more challenging problem to address, and is especially troublesome for AAR studies[46,47] and PMF of eggshell.[87] For mineralized tissues, however, extraction methods can be modified to focus on only mineral-bound and encapsulated proteins. A strong oxidizing agent, such as sodium hypochlorite (NaOCl), can be applied to destroy proteins not encased within mineral,[9,163] leaving behind only intracrystalline proteins.[126] While not necessary in all cases, this aggressive decontamination approach can dramatically improve the proportion of endogenous proteins recovered, even as it reduces the total protein recovery.

2.3.2. Digestion and Digestion-Free Methods. To date, all mass spectrometry studies of ancient proteins have followed a "bottom-up" proteomics approach, meaning that the target of analysis is enzymatically digested peptides rather than intact proteins.[164] Most current protein mass spectrometers are best suited for analyzing peptides in the size range of 6−30 amino acids, and enzyme selection is based on maximizing peptide digests within this range.[165] Trypsin (which cuts C-terminal to arginine and lysine residues), alone or in combination with Lys-C (which reduces missed cleavages at lysine residues) are the most commonly used enzymes for general purpose protein mass spectrometry. Alternative enzymes are also available,[166] and enzymes such as collagenase, elastase, pepsin, chymotrypsin, Glu-C, Lys-N, and ProAnalase have been used in ancient studies to improve coverage of specific proteins or protein regions of interest.[76,167−169] However, proteins characterized by low complexity repetitive domains, as are common in mollusc shell, are difficult to sufficiently digest by enzymatic methods alone and may require additional chemical

cleavage to generate peptides of suitable size for mass spectrometry.[170]

Digestion free methods can be used when the proteins are already broken down into fragment sizes suitable for analysis. This primarily occurs in cases of high diagenetic backbone fragmentation or when proteins are autodigested *in vivo*. A study of 3.8 Ma ostrich eggshell was the first to successfully apply a digestion free method to the recovery of highly degraded struthiocalcin-1 (SCA-1) and struthiocalcin-2 (SCA-2) proteins.[9] More recently, the method has been applied to enamel, including enamel from present-day periodontal patients.[171] Enamel is composed of >98% hydroxyapatite mineral, and a critical step in its maturation is the enzymatic breakdown of the proteins involved in its formation, such as amelogenin.[116] As such, its proteome is already in a fragmented and degraded state during life, with protein fragments in a size range suitable for mass spectrometry. The enamel proteome is also small, comprising only a few major proteins, making it more feasible to analyze and interpret than other proteomes when cleavage positions are variable or unknown. Consequently, digestion-free methods can be applied to enamel proteins and have been used to obtain high quality protein sequences from teeth spanning a wide time range.[13,14,94,172] Such studies of eggshell and enamel are providing our first glimpses of ancient proteins that have been minimally modified by laboratory methods.

True "top-down" proteomics, the measurement and interpretation of mass spectrometry data from intact and native proteins, has not yet been achieved for ancient samples.[127,173] However, ongoing advances in both technology and bioinformatics over the past decade are improving the feasibility of "top-down" approaches.[174]

2.4. Detection by Mass Spectrometry

Once digested peptides or protein fragments have been isolated and purified, they can be analyzed by mass spectrometry. Today, the two main workhorses of paleoproteomics are peptide mass fingerprinting by MALDI-TOF and shotgun proteomics by LC−MS/MS. With respect to cost, time, sensitivity, scale, and scope, each brings different strengths and weaknesses to the study of ancient proteins (Table 1).

2.4.1. MALDI-TOF and Peptide Mass Fingerprinting. Peptide mass fingerprinting (PMF) is a technique used to identify proteins by the masses of the peptides produced following enzymatic digestion. First developed in the 1990s,[175] PMF works best on individual proteins, where ambiguities in peak assignment are minimized, but it can also be applied to proteomes of reliable composition or with one or more dominant proteins, such as collagen in bone or keratins in wool and feathers. PMF was made possible by the development of the soft ionization method matrix-assisted laser desorption/ionization (MALDI) during the late 1980s.[176,177] MALDI represented a major breakthrough in protein chemistry, enabling large, nonvolatile molecules such as small proteins and peptides to be ionized without fragmentation for downstream mass spectrometry. Coupled with a time-of-flight (TOF) analyzer, the MALDI-TOF mass spectrometry system is a robust, simple, and sensitive instrument with a large mass range[175] that is ideally suited for PMF.

To measure protein digests for PMF, acidified peptides are spotted onto a MALDI plate together with a matrix, typically α-cyano-4-hydroxycinnamic acid (CHCA) or 2,5-dihydrox-
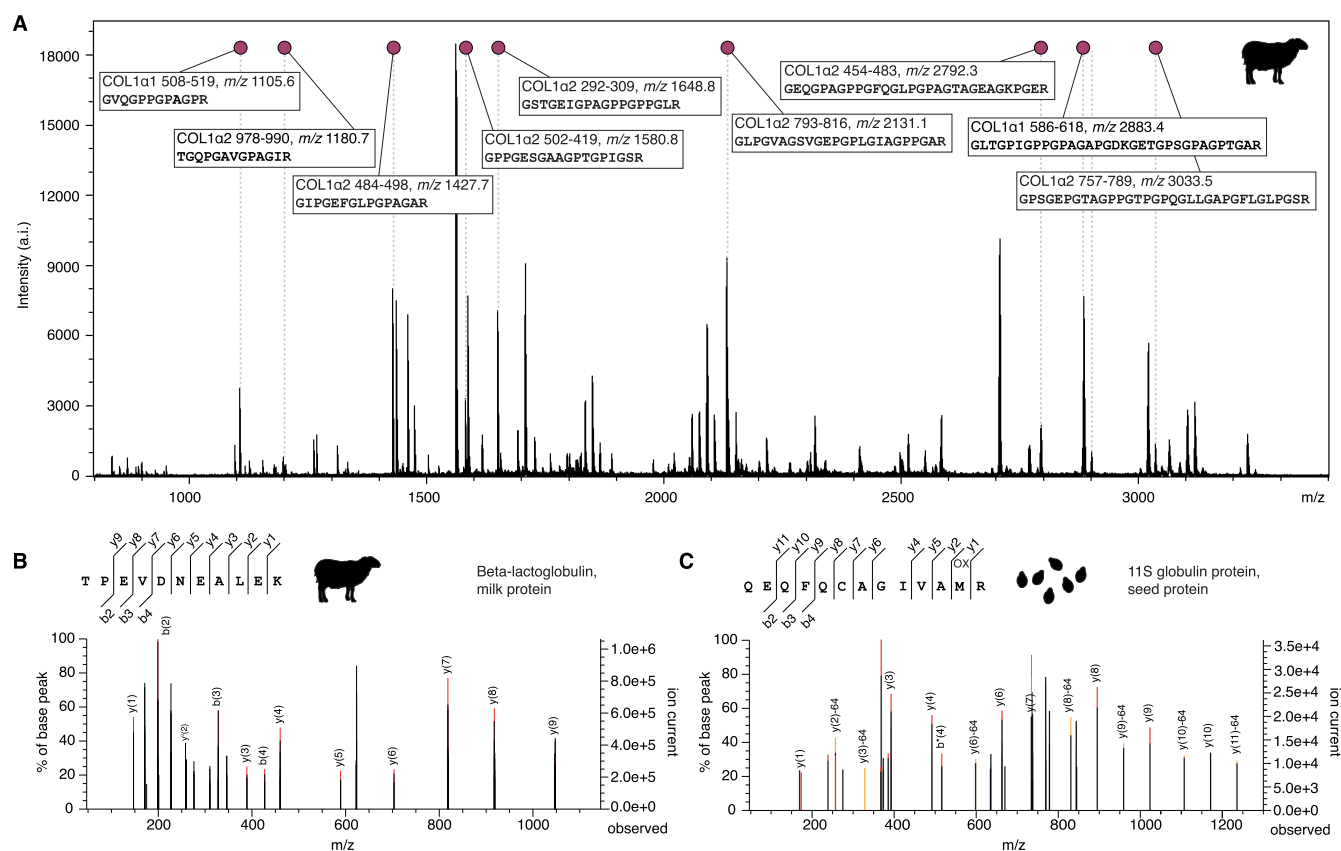
Figure 3. Example MS and MS/MS spectra obtained from archaeological samples. (A) Mass spectrum of sheep (*Ovis*) type I collagen obtained by MALDI-TOF MS from an archaeological small ruminant bone bone at the site of Tepe Yahya, Iran (YTC-248, Peabody Museum No. 986-7-60/22498). (B) Tandem mass spectrum of sheep (*Ovis*) β-lactoglobulin milk protein obtained by nano-HPLC−MS/MS from human dental calculus at the Iron Age pastoralist site of Marinskaya 5, Russia (MKA018). (C) Tandem mass spectrum of sesame seed (*Sesamum*) 11S globulin protein obtained by nano-HPLC−MS/MS from human dental calculus at the Late Bronze Age city of Meggido, Israel (MGD011).[102]

ybenzoic acid (DHB), that cocrystallizes with the peptides. The matrix is then excited with a laser causing the peptides to vaporize and ionize with a + 1 charge. Electromagnets then direct the ions into a time-of-flight tube, where they are measured by a detector. Their time-of-flight (which is related to their kinetic energy and mass) is then converted into a spectrum of mass-to-charge ratios (m/z) vs intensity, and the observed peaks are ready for analysis using a database of protein and contaminant sequences.[72,77,178] Figure 3a shows an example collagen PMF from archaeological animal bone. Informative peaks (markers) originating from the α1 and α2 chains of COL1 are highlighted. Using the principle of parsimony, the nine markers collectively can be used to allow a conclusive assignment to sheep (*Ovis*). Other (nonannotated) peaks visible in the spectrum include matrix peaks, nonmarker collagen peptides, and peptides from keratin contaminants, noncollagenous proteins, and autodigested trypsin. For a more detailed, step-by-step explanation of collagen PMF interpretation, see refs 7 and 77.

Because PMF involves matching a pattern of peaks generated by tryptic peptides, and not sequence determination, it requires access to databases with good taxonomic representation, and identifications are made on the basis of parsimony rather than unique peptide matches. Due to the functional constraints of the protein, sequence variation of COL1 is low, purifying selection is high, and mutational saturation is a challenge for some clades.[179−181] COL1 thus

carries only a weak phylogenetic signal, but if enough marker peptides are sufficiently preserved, taxonomic assignments are generally possible to the family level of birds, the family or subfamily level of mammals, and the genus or species level of fish.[7] For a review of the use of PMF in archaeology, see ref 7.

Despite its limitations, PMF approaches offer several important advantages over other methods. PMF requires very little sample material and is compatible with a number of minimally invasive methods, which will be described below. It does not require specialized facilities, and it utilizes an instrument that is currently widely available at many research institutions and university core facilities. It is also fast and inexpensive, which allows it to be performed at scale and with high throughput. This combination of features makes it a highly flexible method that can successfully support both small-scale budget-restricted projects on specific questions[182,183] as well as large-scale exploratory studies of thousands of samples.[184,185]

2.4.2. LC−MS/MS and Shotgun Proteomics. Tandem mass spectrometry (MS/MS or MS^n), as applied in the context of paleoproteomics, is an approach whereby mass analysis is conducted at least twice while performing a dissociation process in order to characterize peptides in a protein mixture. The first mass scan (MS1) measures the m/z of the ionized peptides (called precursor ions) and selects some for fragmentation by dissociation and further measurement by a second mass scan (MS2) that determines the m/z of the

peptide fragments (fragment ions). Depending on the method of fragmentation, different types of fragment ions are produced. Collision-induced dissociation (CID), which is among the most widely used fragmentation methods, yields mostly b- and y- fragment ions. The MS2 measurement of these fragment ions allow the peptide's amino acid sequence to be inferred with the aid of a database (or even *de novo* under certain conditions[186]), which allows greater confidence in peptide identification than the *m/z* of the precursor ion alone. The MS2 spectra for two peptides recovered from ancient human dental calculus and analyzed by LC−MS/MS are provided in Figure 3. In both cases, a near complete y-ion series was observed, as well as a partial b-ion series, allowing the peptide sequences to be determined with high confidence. The first sequence is a highly specific match to the *β*-lactoglobulin milk protein in sheep (Figure 3b); the second sequence is consistent with the 11S globulin protein of sesame seeds (Figure 3c). The ability to accurately measure both precursor and fragment ions makes MS/MS a powerful technique for identifying ancient proteins.

Although tandem mass spectrometers had been available in various configurations since the late 1960s, it was not until the development of soft ionization methods such as MALDI and electrospray ionization (ESI) in the late 1980s that MS/MS could be applied to proteins.[187] Early uses of MS/MS in paleoproteomics utilized a variety of mass analyzers, including ion traps, quadrupoles, and TOFs[64,74,89] but had relatively low sensitivity and mass accuracy, which limited the number of proteins that could be identified. The commercial introduction of the Orbitrap mass analyzer in the mid-2000s and subsequent hybrid systems marked a major improvement in protein mass spectrometry[188] and dramatically improved the detection and identification of proteins in low biomass, complex mixtures, which are characteristic of ancient samples. Gains in ancient protein identifications were enormous from the first commercial model, the Thermo LTQ Orbitrap, which identified three collagen proteins in a mastodon bone in 2007,[189] to the subsequent Thermo LTQ-Orbitrap Velos, which enabled the identification of a proteome of more than 100 proteins in a mammoth bone in 2012,[91] and then the Thermo Q-Exactive Hybrid Quadrupole Orbitrap, which allowed the characterization of a metaproteome comprising hundreds of proteins in human dental calculus in 2014.[92] Current MS/MS systems used in paleoproteomics are even more powerful and typically consist of an ultrahigh perform-ance liquid chromatography (UHPLC) system coupled to a nano-ESI that interfaces with a high performance (high resolution fast duty cycle) mass spectrometer. Greater chromatographic separation, fractionation of samples, and/or the use of ion mobility[190,191] can further enhance resolution, and alternative ionization methods, such as desorption electrospray ionization (DESI) and liquid extraction surface analysis (LESA),[192−195] offer additional capabilities, including ambient ionization and *in situ* analysis; however, these approaches have not yet been extensively explored in paleoproteomics.

Tandem mass spectrometry is well-suited for the analysis of ancient proteins from diverse sample types. It can identify high and low abundance proteins in complex mixtures, and it does not require reliable - or even known - protein composition prior to analysis. Because it involves the simultaneous analysis of many proteins, it can be used to achieve higher taxonomic resolution than PMF, which is particularly important for

resolving vertebrate[14,21,196] phylogenies. Moreover, it can also be used to identify protein variants, PTMs, and diagenetic alterations. These features make it ideal for discovery proteomics applications, such as phylogenetic analysis of extinct hominids,[13,172] taxonomic identification of worked shell,[197] determination of unknown binders in artwork[137] and the identification of dietary proteins in pottery crusts[99] and dental calculus.[101,102] Beyond shotgun approaches, MS/MS can also be used to some extent to target ancient proteins of interest using multiple reaction monitoring (MRM; also known as selected reaction monitoring, SRM)[198−201] and parallel reaction monitoring (PRM).[202] Although current MS/MS approaches largely rely on data-dependent acquisition (DDA) for precursor ion selection, which maximizes the success of peptide sequence determination but limits the method's reproducibility and quantitative potential, data-independent acquisition (DIA) approaches[203,204] are in now development for ancient protein analysis. DIA offers the potential to extend the dynamic range of MS/MS by generating data from more peptides, and especially lower abundance peptides, while also improving reproducibility and quantification. A DIA-based approach has recently been integrated into a new paleoproteomics workflow known as species by proteome investigation (SPIN), which enables rapid mammalian species assignment using LC−MS/MS.[205] Although currently limited in scope due to its computational complexity, further improvements in DIA development, improved databases, and the application of machine learning may soon allow DIA to become more mainstream in ancient protein studies.

The major downside of MS/MS in ancient proteomics is its significant infrastructure needs, time, and cost. Samples for MS/MS should be prepared in a dedicated ancient biomolecules laboratory, in part because the higher sensitivity of the instruments and the discovery nature of the research makes distinguishing ancient proteins from contamination more difficult. Highly specialized and expensive mass spectrometers are required that may not be widely available at local core facilities, and the instrument time per sample is high (an hour or more), limiting daily throughput. Currently, the costs of MS/MS are typically 30−50 times higher on a per sample basis than PMF, although new specialized applications, such as SPIN, are faster and more affordable. Despite its difficulties, the power and performance of LC−MS/MS, and most importantly its ability to provide sequence data, make it a highly valuable—and even indispensable—technique to answer many paleoproteomic questions.

2.5. Analysis and Interpretation of Data

With the exception of largely experimental work still in development,[186] nearly all mass spectrometry data analysis methods relevant for ancient proteins rely on the use of specialized software, protein sequence or peptide marker databases, and the selection of priors.

2.5.1. MALDI-TOF and ZooMS.
Analysis of PMF data for the purpose of taxonomic identification is most frequently performed manually through the visualization of spectra using FlexAnalysis (Bruker Daltonics) or mMass[206] software. Unfortunately there is no centralized database or public repository of PMF spectra or markers at present, and consequently peptide markers (peaks that have been empirically demonstrated to be taxonomically informative) must be retrieved from literature searches. Taxonomic

identifications are made by applying the principles of parsimony to the combination of markers observed.[77] Depending on the potential species present, additional tools have been recently proposed to help with identification based upon machine learning, hierarchical clustering, principal components analysis (PCA), and theoretical spectra matching.[207,208] During spectra interpretation, it is essential to consider the possibility of mixed or composite proteome representation (e.g., in the case of glues), and potential contamination must also be taken into account.

2.5.2. LC−MS/MS, Protein Identification, and *De Novo* Sequencing. The analysis of MS/MS data is usually conducted with the aid of software that matches precursor ion masses and individual MS2 spectra against a set of theoretical masses calculated from a protein database and preselected priors, producing a peptide match score, as well as other metrics. This approach requires the input of: (1) a database, (2) instrument parameters, and (3) search priors. Some software also integrates *de novo* sequencing, machine learning algorithms, or other alternative workflows, which can reduce the reliance on databases and enable the characterization of novel sequences. The most commonly used software for paleoproteomics are MASCOT,[209] MaxQuant,[210] SEQUEST,[211] PEAKS,[212] and Byonic.[213] Additional software, such as Scaffold,[214] can be used to further authenticate and filter the results by protein and peptide probability and by false discovery rate (FDR), and peptide identifications are typically manually validated by searching them against the NCBI nr database using BLASTp to ensure specificity.

Although reference proteins can be directly sequenced, the vast majority of protein sequence data available in major protein databases derives from genetic coding sequences (CDS) submitted to NCBI (GenBank), EMBL-EBI (EMBL-Bank), and DDJB, the three major public nucleic acid databases that together form the International Nucleotide Sequence Database Collaboration (INSDC). Other sources of genome-derived annotated protein sequences include NCBI RefSeq[215,216] and Ensembl Genomes,[217,218] as well as WormBase[219] and ParaSite[220] for parasitic nematodes, and VectorBase[221] for pathogen vector genomes. UniProtKB, the world's largest public repository of protein information, aggregates data from the INSDC and releases it in two databases: (1) SwissProt, which contains manually annotated and reviewed sequences; and (2) TrEMBL, which consists of all remaining nonreviewed, automatically annotated sequences. Specific databases are also available for individual species proteomes (e.g., UniProt Proteomes, NCBI RefSeq), and custom databases can be created using curated lists of genomes, such as the Human Oral Microbiome Database (HOMD),[222] or metagenomes, such as NCBI env_nr. In addition, specialty databases for identifying common laboratory contaminants, such as the common Repository of Adventitious Proteins (cRAP; https://www.thegpm.org/crap/) and databases for assessing preservation in specific ancient sample types, such as dental calculus,[223] have also been developed. These databases contain different numbers of sequences, varying levels of metadata, and different database biases.[224] Decoy databases, or other integrated target-decoy search procedures, are included during analysis in order to calculate FDR.[225] Care should be taken to ensure that the selected databases are appropriate for the sample type. For example, studies of microbial substrates, such as dental calculus and paleofeces, or tissues that have been degraded by

environmental bacteria, such as skeletal and mummified tissues, should include microbial proteomes in database searches in order to ensure that microbial proteins are not better matches for spectra putatively assigned to dietary or host-derived proteins. Likewise, investigations of pathogens (e.g., *Mycobacterium tuberculosis*) should also include protein sequences from related microbial taxa (e.g., soil *Mycobacteria*) in order to ensure taxonomic specificity. Discovery-based proteomics, especially when applied to complex metaproteomes, is particularly sensitive to database selection, and overly restrictive databases should be avoided in order to mitigate false positives.

Once one or more databases have been selected, instrument parameters and search priors are required. Instrument parameters correspond to the particular mass spectrometer used for the analysis, and include information about peptide ionization and MS1 and MS2 mass accuracies. Search priors relate to the sample itself, and consist of information about the enzymatic digestion, isotopic composition, and anticipated chemical modification of the peptides. This includes specifying the digestive enzyme (e.g., trypsin), which should match the experimental enzyme used, as well as the assumed fidelity, making allowances for missed cleavages (typically 1−3). The number of anticipated ^{13}C atoms is typically preselected, and fixed and variable chemical modifications resulting in mass changes are also specified as priors. These include any intentional modifications introduced during extraction (e.g., carbamidomethylation of cysteine), common biological PTMs (such as oxidation and phosphorylation), and diagenetic modifications (such as deamidation and glycosylation).

While the ability to match against every known protein sequence and all possible chemical modifications would be ideal for ancient samples, computational effort scales linearly with the number of sequences in the database and exponentially when relaxing chemical modification and digestion parameters. Therefore, choices must be made to limit the search space to allow reasonable computational efforts. For example, *in vivo* and diagenetic backbone cleavage can be accounted for in searches by selecting enzyme options such as "semi-trypsin", but this increases the search space of the algorithm. Likewise, searches can also be conducted in "error tolerant" mode, which allows for amino acid substitutions and unspecified chemical modifications, but this increases the FDR and can reduce the number of successfully identified ancient proteins. For ancient samples, error tolerant searches are generally reserved for assessing the range of chemical modifications in a sample prior to further analysis (e.g., refs 92 and 102) or for determining novel sequences in already well characterized proteins (e.g., collagen[226]) or proteomes (e.g., enamel[13,172]).

In phylogenetic studies of taxa for which the exact protein sequence is unknown, software is available for *de novo* sequencing.[227] This is especially valuable for determining sequences in extinct species for which genomic data cannot be obtained[20,172,226,228] or authenticating PMF markers.[73,182,229] A variety of software has been used to perform *de novo* sequencing on ancient proteins, including PEAKS, Byonic, and MaxNovo. Algorithmic differences between the different software influence the accuracy of identification. The taxonomy of tryptic peptides identified by *de novo* sequencing can be inferred using tools such as UniPept.[230]

After a search is completed, potential identifications are evaluated on the basis of a number of metrics, including

peptide match scores, the number of assigned peptides for a given protein, and the peptide and protein FDRs. Taxonomic assignment for high scoring identifications should be further validated for specificity using additional sequence alignment tools, a diverse database, and manual validation of critical spectra. It is essential to take into account potential database and search biases when interpreting results, especially when conducting discovery-based proteomics studies. Great care must be taken in selecting parameters to obtain optimal and accurate results,[231,232] and improbable results[233,234] should be subject to further scrutiny. Recommendations for minimum standards of authentication and validation are described in ref 4.

2.5.3. Taxonomic Discrimination. Once peptides are identified to proteins, their sequences can be used to infer taxonomy and to discriminate between related taxa.[20,73,86,88,235] The taxonomic resolution of a given set of peptide sequences can vary widely, however, depending on the evolutionary history of the protein and the specific evolutionary forces acting on it. Overall, protein sequences provide less taxonomic resolution than DNA sequences, but proteins can persist millions of years longer than DNA, they are biologically present in much higher amounts than DNA, and they are found even in acellular tissues (e.g., enamel, eggshell). As such, they are our most valuable form of molecular sequence data for providing successful and reliable taxonomic identifications in deep time fossils and in the study of processed and manufactured objects that have undergone activities that are destructive to nucleic acids (e.g., leather tanning, liming of parchment).

Overall, sequence change is generally more constrained and occurs more slowly in proteins than in DNA. This is because, unlike DNA, where noncoding regions and redundancy in the genetic code allow for mutations independent of selective pressures, amino acid changes directly affect the protein and—depending on the chemical properties of the altered amino acid and the location of the substitution within the peptide—can strongly influence the protein's structure and function.[236,237] In addition, the synthesis or acquisition of specific amino acids for incorporation into proteins can have different metabolic costs depending on the organism and the environment.[238] This means that almost all proteins are under some level of selective pressure, and substitutions between amino acids with similar chemistry generally occur at a higher rate than amino acids with different chemistry.[239,240] Consequently, there is a higher probability for convergent, parallel, and back substitutions in proteins, which can result in distantly related species sharing the same substitutions (homoplasy)—a particular problem for proteins under high functional constraint.[241−244]

Protein sequence conservation among closely related taxa determines the level to which taxonomic discrimination is possible. In some cases, taxonomic discrimination is not possible because there are no sequence differences between a set of taxa for a given protein of interest. For example, wild (*Ovis ammon*, *Ovis gmelini*) and domesticated (*Ovis aries*) sheep have identical milk β-lactoglobulin protein sequences, and thus, taxonomy derived from this protein cannot be assigned at a level lower than that of genus (*Ovis*). Even in cases where sequence differences do occur, taxonomic discrimination may fall short of theoretical predictions because not all variant sites fall within peptides that are likely to be observed using mass spectrometry, either because the corresponding tryptic peptides are too long or too short or

because they are too hydrophobic. For example, the COL1a2 amino acid sequences of domesticated horse and donkey differ by 4 out of 1038 residues in the mature protein. Theoretically, discrimination between the two species should be possible on the basis of these four residues, but in practice tryptic digests of this protein result in all four of these taxonomically informative amino acids falling in peptides that are unlikely to be detected by mass spectrometry. As a result, horses, donkeys, and mules (horse/donkey hybrids) cannot be distinguished using standard ZooMS techniques. In contrast, sheep and goat COL1a2 proteins, despite being even more similar (differing by only 2 out of 1038 residues), are generally distinguishable by ZooMS because their taxonomically informative residues fall on a tryptic peptide that is frequently observed by MALDI-TOF mass spectrometry.[73,78,245]

Taphonomic alterations and digestion efficiency can also further influence taxonomic specificity. Within the β-lactoglobulin protein, for example, the tryptic peptide TPEVD-(D/N/K)EALEK is the most frequently recovered peptide containing a taxonomically variant site.[4] However, the residue that distinguishes cattle (D) from sheep (N) is unreliable in archaeological samples because a cattle aspartic acid (D) cannot be distinguished from a taphonomically deamidated asparagine (N). Thus, peptides bearing the aspartic acid residue must be provisionally assigned as cattle/sheep. Moreover, the lysine (K) residue in this peptide that distinguishes goats is also a tryptic cut site (which cuts the peptide into fragments too short for MS detection), and thus, a taxonomic assignment of goat can only be made if there is a missed tryptic cleavage in this peptide. Such complications must be factored into strategies for taxonomic discrimination of archaeological proteins, and as a result, the taxonomic resolution of ancient protein data in practice is often lower than would be predicted from protein sequence alignments alone.

In some cases, additional metadata, such as the time period or location from which the sample was obtained, can provide information allowing a greater degree of taxonomic specificity to be inferred. For example, *Ovis* β-lactoglobulin sequences obtained from locations outside the range of wild *Ovis* species, such as in colonial-era Americas, can be reasonably assumed to have originated from domesticated *Ovis aries*. Likewise, Holocene-era *Bos* sequences obtained from European Neolithic sites postdating the decline of aurochs (*Bos primigenius*) and outside the range of zebu (*Bos indicus*) and yaks (*Bos mutus*, *Bos grunniens*) can be reasonably assigned to domesticated cattle (*Bos taurus*). Such context-based inferences, however, must be applied with care, especially in places and periods where multiple species may have been present or where former species ranges are not well-known.

Finally, since mass spectrometry recovers peptides and not entire proteins, and these peptides have varying levels of taxonomic specificity, taxonomic assignments are typically made on the basis of parsimony, with the assumption that the peptides derive from the fewest number of species possible. In the case of endogenous tissues, such as bone, peptide sets for proteins such as collagen are assumed to derive from a single organism. However, the assumption that all peptides derive from a single organism does not hold in the case of manufactured or mixed proteomes, such as collagenous glues, dental calculus, or pottery food crusts. In these cases, taxonomic identification is undertaken by examining the full range of peptide sequences obtained from many proteins, and

also taking into account biogeographical and archaeologically relevant prior information in order to make the most reasonable taxonomic identification(s) from the available data.

3. APPLICATIONS IN PALEOPROTEOMICS

Although ancient proteins are studied in a wide variety of ways in the fields of archaeology, cultural heritage, and paleontology (for a review see ref 5), most applications can be grouped into one of three broad categories based the nature and composition of the sample: (1) proteins, where a single protein is the primary target of analysis; (2) proteomes, where groups of endogenous proteins inherent to a tissue or substrate are studied (e.g., bone, enamel, shell); and (3) metaproteomes, where protein and proteome mixtures of diverse biological or manufactured origin are analyzed (e.g., dental calculus, paleofeces, pottery crusts, artist materials). Here, we describe each application in turn and highlight its uses, strengths, and challenges.

3.1. Proteins

A large proportion of ancient protein research has been dedicated to the detection and taxonomic assignment of highly abundant proteins that form the dominant structural components of tissues. PMF by MALDI-TOF is particularly suited to this situation, but when preservation is poor or when higher resolution sequence data is needed, more powerful tandem mass spectrometry approaches, such as MALDI-TOF/TOF or LC−MS/MS, can be applied. The most frequently analyzed ancient protein is COL1, a robust structural protein[246−248] that is capable of surviving more than 3 million years under ideal conditions.[10] Other proteins of interest include keratins[84] to identify wool, horn, hair, feathers, turtle shell, and baleen, and fibroin,[249] the dominant protein component of silk. In addition, the enamel protein amelogenin, which is important for tooth formation, can be used to determine the genetic sex of some mammal species.[94,250]

While mineralized tissues, such as bone, enamel, and shell, are frequently the target of ancient protein analysis due to their durability, some proteinaceous soft tissues are also suitable for analysis. Collagens, keratins, and fibronins are collectively the major components of textiles and parchments produced from primary and secondary animal products, such as hide, skins, leather, wool, fur, felt, and silk. These materials are often culturally important, but ephemeral and underrepresented in the archaeological record,[251] although their preservation can be enhanced by contact with antimicrobial metals such as copper. Paleoproteomic methods can increase their visibility by improving taxonomic identifications and identifying trace remains.

3.1.1. Collagens: Bone, Dentine, Antler, Ivory, Parchment, Leather, Gut, and Scales.
ZooMS analysis of COL1 is the most frequently conducted type of paleoproteomic analysis,[178] and it can be conducted on almost any collagenous tissue, including mineralized tissues such as bone, dentine, antler, ivory, and horn core,[72,252,253] as well as nonmineralized tissues such as skin, parchment, leather, gut, scales, and other soft tissues.[156,160,254] It is especially useful for identifying material that has lost its diagnostic features, such as worked bone[159] and bone fragments,[185] and it can be used to screen large numbers of nondiagnostic fragments for species of interest.[184] It has also been proposed as one possible screening method for assessing collagen preservation prior to radio-

carbon dating,[255] although ZooMS requires less collagen than either radiocarbon dating or stable isotope analysis.[139]

ZooMS identification using MALDI-TOF is performed using a database of taxonomically informative marker peaks, and it is important that the sequences of newly developed markers are first verified as authentic collagen sequences with taxonomically informative amino acid substitutions using MALDI-TOF/TOF or LC−MS/MS,[79,178] although there are cases where this has not been feasible.[182,256] Large mammals, particularly European species, make up a large proportion of the published markers, e.g.,[72,257,258] but markers for other taxonomic groups are increasingly being developed, including for non-European large mammals,[79,252,259−261] rodents,[262,263] bats,[262] cetaceans,[80,264] marsupials,[265] birds,[77,82] fish,[81,182,256] amphibians,[266] and reptiles.[83,262,266,267]

Over the past decade, ZooMS has been used to answer a wide range of cultural heritage, archaeological, ecological, and paleontological questions. For example, ZooMS has been used to study the manufacture of worked bones, artifacts, and cultural heritage materials[159,245,252,254,259,260,268−274] and to better characterize archaeological faunal assemblages and past human−animal relationships.[80,82,185,257,263,275] It has been used to better define past domestic animal management strategies,[73,261,276−279] document the introduction of commensal species associated with human activities,[280−284] and identify the exploitation of wild species.[79,83,285,286] It has contributed to the reconstruction of past ecologies[182,256,262,265,267,287−289] and to the study of extinct megafauna.[196,265] ZooMS has also been notably used as a low-cost, high-throughput screening tool of bone fragments in large Pleistocene cave sequences, leading to the discovery of otherwise nondiagnostic hominid remains,[144,184,257,290] including the offspring of a Neanderthal mother and Denisovan father.[291,292] Finally, because ZooMS can be performed using minimally invasive sampling techniques,[154,156,157,159] it has proven a breakthrough technology in the emerging field of biocodicology,[293] the multidisciplinary analysis of parchment manuscripts, codices, and other historic documents.[156,208,294−303]

3.1.2. Keratins and Corneous β-Proteins: Wool, Hair, Feathers, Baleen, and Turtle Shell.
α-Keratins and corneous β-proteins (CBPs, formerly β-keratins) are two of the most important structural protein classes in vertebrates after collagens, and they are the major components of hair/fur, nails/claws, horns, hooves, feathers, beaks, turtle shells, quills, and baleen. They are also present, along with collagen, in skin.[304−306] Like collagens, dozens of individual keratin and CBP proteins have been characterized.[307,308] They can also be taxonomically distinguished by PMF using MALDI-TOF.[309] Unlike collagen-producing cells, keratin-producing cells die after producing keratins and CBPs, and consequently keratinous tissues are not living and do not remodel.[304] Keratinous tissues do not mineralize, and even the hardest CBP tissues (e.g., turtle shells, beaks, claws, etc.) rarely contain minerals and therefore are more subject to degradation than mineralized proteins.[310] However, under favorable preservation conditions keratins and CBPs can preserve due their hydrophobicity and resistance to many proteases,[311] and they are also found archaeologically when embedded in mineral matrix produced by the degradation of nearby metal items (e.g., weapons, crowns, pins, buckles).[104,310]

MALDI-TOF marker peptides to identify furs and textiles were originally developed using PCA methods and have been subsequently verified using LC−MS/MS or MALDI-TOF/

TOF, providing genus level resolution for some groups of mammals[104,109,251,309,312−317] and whale baleen.[318] The only taxonomic group with available CBP markers providing genus level resolution is sea turtles.[310,319] The use of immunological assays to detect wool on metal artifacts and from textile imprints in soils is currently being investigated.[320,321] Increasing understanding of keratin and CBP diversity, for example differences in sheep wool pigmentation and curl/ crimp related to domestication, selective breeding, and diet[322,323] and keratin texture variation associated with body location, disease, and age in humans,[324,325] may allow more information to be gleaned than just taxonomic classification.

3.1.3. Fibroin: Silk. Silks are structural proteins, composed of highly repetitive β-sheet motifs interspersed with flexible domains. Raw silkworm silk is composed primarily of two proteins, fibroin and sericin, with several additional proteinases and functional proteins making up a minor component. During textile production, the nonfibroin proteins are removed, leaving fibroin as the dominant protein in archaeological silk.[251,326] While silks in modern and historic periods all derive from a single domesticated silkworm species from China (*Bombyx mori*), a number of other insects can also produce silk, including other wild silkworms and moths (especially *Bombyx mandarina*, *Samia cynthia*, *Antheraea* sp., and *Philosamia* sp.) and spiders (e.g., *Nephila clavipes* and *Araneus diadematus*), and use of these silks has been documented historically and archaeologically in Asia, India, Europe, North America, and Australia.[251,327−330] Silks from *B. mori* and *Antheraea pernyi* can be distinguished proteomically by LC− MS/MS and immunological assays.[251,331] This has allowed silk textiles and their species of origin to be identified from sediments with textile imprints[249,332] as well as trace amounts of textile in contact with metal artifacts.[333,334] Further work on the characterization of fibronins from other species has the potential to greatly improve the understanding of silk production and trade and silkworm domestication.

3.1.4. Amelogenin: Sex Typing of Humans and Other Mammals. Amelogenin (gene *AMEL*) plays an important role in enamel formation and mineralization in newly secreted enamel. During enamel maturation, amelogenin is cleaved by proteases and trapped within the enamel matrix.[116,335] In monotremes, marsupials, and nonmammalian species, *AMEL* is an autosomal gene, while in eutherian mammals it is located on the sex chromosomes.[250,336] In most mammals, only *AMELX* is functional, and AMELY expression is low. In some species, there are no sequence differences between *AMELX* and *AMELY*. Other species have lost *AMELY* entirely. However, for species with sequence differences between *AMELX* and *AMELY* and where AMELY is expressed, protein analysis can allow sex descrimination.[336,337] These species include humans,[338,339] cattle,[340] bison,[341] sheep,[342] goats,[343] deer,[342] pigs,[344] horses,[345] and bears.[346]

Before the widespread application of LC−MS/MS to archaeological samples, attempts to use MALDI-TOF/TOF to identify both *AMELX* and *AMELY* for sex determination were met with minimal success.[347,348] Subsequent efforts using LC−MS/MS have proven much more productive for sex determination in ancient humans[94,171,198,202,349−354] and archaic hominids.[172] Sex determination for other species has been limited to date, but includes extinct species of rhinoceros (*Stephanorhinus* sp.[14]), probocideans (*Notiomastodon platensis*), and rodents (*Myocastor coypus*[355]). Although there are concerns that the method could produce inaccurate results due

to the presence of low frequency of *AMELY* deletion variants in some populations,[356,357] sexing using amelogenin is generally a robust technique, and it is an important biomolecular tool in cases where mophological sex determination is not possible, especially for incomplete skeletons and juveniles,[94] or when ancient DNA analysis is unsuccessful or infeasible.[354]

3.2. Proteomes

The first ancient proteome, from mammoth bone, was published in 2012.[91] Consisting of more than 100 proteins, it was a singular achievement and marked the field's transition from the study of ancient proteins to true paleoproteomics. Proteomes are the suite of proteins present in a given tissue, and while the genome of an organism remains relatively constant throughout the body and lifetime, the proteome can differ substantially. Proteomes consist of protein mixtures, which often have a high degree of complexity and a wide dynamic range of protein expression.[358,359] LC−MS/MS approaches are necessary to profile proteomes, but comprehensive proteome characterization is still highly challenging even for modern proteomes, let alone ancient ones. Nevertheless, significant progress has been made, particularly toward characterizing the proteomes of ancient bone, enamel, eggshell, and mollusc shell, with mummified tissues being less explored. Although proteome data can be used to provide taxonomic assignments, they are also capable of addressing much more complex questions regarding phylogenetic relationships, health, and aging and development. Below, we describe the current state of research on ancient proteomes and explore their applications.

3.2.1. Bone and Dentine. Bone and dentine have similar developmental origins and correspondingly similar, but not identical, proteomes.[141,360] While they are both dominated by the structural protein COL1, they also contain a diverse array of hundreds of other collagens and non collagenous proteins (NCPs).[100,361−363] Many proteins are shared between bone and dentine, including interstitial fibrillar collagens (e.g., types I, III, and V), as well as proteins that support collagen fibril organization (e.g., lumican, LUM) and facilitate biomineralization (e.g., biglycan, BGN; fetuin-A, AHSG). Other proteins are more tissue specific, such as periostin (POSTN), which is disproportionately expressed in bone periosteum, chondroadherin (CHAD), a cartilage-associated protein expressed on bone articular surfaces, and asporin (ASPN), a protein that facilitates tooth attachment to the periodontal ligament. Plasma proteins and blood clotting proteins are also typically present in bone and dentine proteomes, including prothrombin (F2), coagulation factor IX (F9), and coagulation factor X (F10). To date, ancient bone proteomes have been studied in humans,[96,168,258,364,365] mammoths,[91,366] moas,[128] cattle,[121,132] horses,[12,168] turkeys,[367] rabbits,[367] squirrels,[367] extinct rhinoceros,[14] and ancient dentine proteomes have been studied in humans[92,141,365] and extinct rhinoceros.[14]

Because many of the NCPs and some of the collagens have higher mutation rates than COL1, they are better targets for phylogenetic reconstructions, especially between closely related species.[258,368] Analyses of bone and dentine proteomes have successfully aided in resolving the phylogenetic relationships of extinct megafauna,[14,20,21,196] including archaic hominins.[258,365] In addition, bone proteome-scale analyses using new high-throughput LC−MS/MS workflows, such as

SPIN, show great promise for enabling genus and species-level taxonomic assignments of nondiagnostic bone.[205]

Proteome-level analyses can also be used to detect altered bone proteomes, such as those that occur in association with changes in activity levels,[369] health and disease,[370] and age.[371] Chronic bone infections, such as osteomyelitis, and bone cancers are known to alter the bone proteome,[372–375] and a number of studies have attempted to diagnose archaeological cases of cancers based on proteomic evidence,[96,376] although the results so far have been mostly qualitative and nonspecific. Changes in bone proteome composition and chemical modifications also occur with advancing biological age across multiple tissues, including bone marrow,[371,377] and composition changes and modifications correlated with biological age and post-mortem interval have been observed in the proteomes from modern and archaeological skeletal material.[141,258,364,378–382] However, bone is one of the poorer characterized tissues in the Human Proteome Organization (HUPO) Human Proteome Project[383] and proteome changes observed in archaeological bone samples could be the result of diagenetic factors in addition to biological age, physiological stress, life history traits, and disease. More work is needed to assess the reliability of chemical modifications and proteome composition as signals in archaeological samples.

3.2.2. Enamel. Enamel is the hardest tissue in the vertebrate body.[384] Composed mostly of hydroxyapatite crystals, the amount of organic matrix in mature enamel is very small, making up less than 2% of its total mass.[385] Enamel has a small proteome, consisting of only five abundant proteins: amelogenins (AMELX and AMELY), ameloblastin (AMBN), amelotin (AMTN), enamelin (ENAM), and odontogeneic ameloblast-associated protein (ODAM). Of these, amelogenin is the most abundant and best characterized, comprising 90% of all enamel protein.[385] Other proteins, including keratins (e.g., KRT75),[386,387] as well as collagens and blood proteins,[116] have also been detected in enamel at trace levels, but the latter likely represent contamination from dentine. Enamel proteins are degraded *in vivo* by the matrix proteases enamelysin (MMP20) and kallikrein related peptidase 4 (KLK4), resulting in only enzymatically cleaved proteins being present in mature enamel.[384,385] These degraded proteins are concentrated in the enamel tufts at the dentine–enamel junction (DEJ) and along hydroxyapatite rod sheaths.[386] Although early attempts to directly analyze enamel proteins from modern and archaeological teeth typically only detected amelogenin,[348,388] methodological improvements have increasingly allowed the recovery of a richer enamel proteome.[14,94,171]

Due to the protective nature of enamel in comparison to bone, enamel proteins are some of the longest surviving proteins in the vertebrate body.[171] Although the enamel proteome is small and carries less phylogenetic resolution than other tissues, such as bone, ancient enamel proteins have nevertheless been used to successfully determine the phylogenetic relationships of extinct species,[13,14,172,355] and enamel proteins are especially valuable in cases when DNA or other proteins are unlikely to survive.[389]

3.2.3. Avian Eggshells. Like enamel, the shell biomineral matrix provides the potential for better protection from water and thus degradation of proteins than bone.[9] As such, eggshell holds great promise for surviving into deep time, and eggshell currently holds the record for the oldest successfully determined and independently verified peptide sequence—from a 3.8 Ma ostrich eggshell in East Africa.[9]

Avian eggshells mineralize around the egg membrane in a multistage process mediated by the organic matrix proteins.[390,391] Although ∼500−1000 proteins have been identified using proteomic methods in fresh eggshell,[392,393] considerably fewer are typically identified in archaeological samples.[86] Prior to analysis, archaeological eggshells are typically treated using bleach to remove intercrystalline components, leaving only intracrystalline proteins for analysis. LC−MS/MS analysis of eggshell has recovered a wide range of eggshell proteins, notably ovocleidins, ovocalyxins, ovalbumin, struthiocalcin, rheacalcin, ansocalcin, ovomucoid, ovotransferrin, ovostatin, and mucins, among others.[86]

Further characterization of eggshell proteomes using both MALDI-TOF and LC−MS/MS[86,87] have allowed the development of a large number of marker peptides that can reliably distinguish common bird taxa (most importantly chicken, duck and goose) using PMF.[394–397] PMF is a rapid and inexpensive method for identifying archaeological eggshell, but due to the complexity of the eggshell proteome and the limited number of characterized species, not all eggshell can be identified using this method. When MALDI-TOF analysis is unable to taxonomically identify the species, LC−MS/MS analysis can be conducted on those samples that have distinct MALDI-TOF spectra in order to facilitate identification.[398,399] Success will increase as more bird genomes are sequenced and deposited in genomic and proteomic databases.[392]

Eggshell proteomes are powerful means to increase the taxonomic resolution of archaeological eggshells and allow more nuanced interpretation of the interactions between humans and birds at archaeological sites. For example, because diet and age influences eggshell strength and quality,[400] future studies of archaeological eggshell may be able to provide insights regarding the early husbandry and feeding practices of domesticated birds, such as turkeys,[401] as well as captive wild birds subjected to intensive breeding programs, such as scarlet macaws.[402,403]

3.2.4. Mollusc Shells. Molluscs have evolved a biomineralized exoskeleton, or shell, that provides support, protection, and defense and also serves as a reserve of calcium ions.[404] Important to proper shell formation are the shell matrix proteins that can persist in the intracrystalline matrix of biomineralized shells, where they are protected from contamination and degradation into deep time.[405–409] Shell matrix protein sequences are diverse and therefore have the potential for providing genus or even species-level taxonomic resolution.[404,410,411] Recent protocols have been developed that can successfully extract shell matrix proteins ("shellomes") from worked and unworked shells for analysis by MALDI-TOF and LC−MS/MS.[88,197,412] Although identification is currently limited by the lack of available reference data,[411] "palaeoshellomic" analysis holds great potential for revealing humans' longstanding relationship with shell and shellworking,[197] which stretches back more than 100,000 years,[413] and will aid in the better understanding and resolution of amino acid geochronology.[126]

3.2.5. Mummified Remains. Although mummification is relatively uncommon, it does occur under a variety of artificial and natural conditions, providing a rare opportunity to study ancient soft tissues.[414] The most frequently preserved and analyzed soft tissue is skin. Although most proteomic studies of skin (and treated skin products, such as leather and

parchment) have focused only on collagens, skin is a complex tissue with a diverse proteome that also includes a large number of less abundant proteins.[415,416]

Using LC−MS/MS, much of the ancient skin proteome can be accessed, and studies of human skin to date have focused on characterizing proteome preservation in artificially[417] and naturally[418] mummified skin, with an emphasis on documenting proteins associated with innate immunity. Studies of animal skins have gone beyond taxonomic assignment to further identify fetal proteins differentially expressed in the months just before and after birth (e.g., RPN2, HBBF, HSP90A), allowing the identification of calf skin and an estimation of the age of death of the animal.[106,156] However, despite the large number of questions that can be potentially addressed by analyzing the proteomes of ancient skins, hides, and parchments, the study of ancient skins has continued to focus mainly on collagen. Studies of ancient skins are challenged by the unavoidable ubiquity of modern human skin and hair proteins, and the interpretation of proteins from ancient skins requires a good understanding of their excavation and curation history in order to account for contamination. Additionally, until very recently, sampling of ancient skins required destructive techniques that were strongly discouraged by museums and archives. Recent successes with minimally invasive EVA films on mummified tissues[419] and PVC rubbings on parchments[156] are changing the landscape of ancient skin studies and will likely increase the number of available samples that can be analyzed.[419]

In addition to skin, other mummified soft tissues and organs have also been documented and occasionally analyzed using proteomic techniques, including muscle[417] and stomach tissue.[420] More surprising are the hundreds of preserved brains that have been characterized,[421] with likely thousands more being preserved around the world.[422] At the time of writing only three have undergone proteomic analysis,[422−425] with highly variable numbers of proteins recovered. The success of brain preservation is likely due to the formation of protein aggregates that provide protection against degradation.[422,424] While soft tissues still remain relatively rare in comparison to other sample types, protein markers for health, age, and life history could potentially be developed.

3.2.6. Plant Macroremains.
Plant macroremains (e.g., seeds, fruits, wood) can preserve under extraordinary conditions, such as in waterlogged or charred states, or in cases of extreme aridity or cold.[426] Although modern seeds contain proteins numbering in the hundreds to low thousands,[427,428] protein recovery reported from ancient seeds has been much lower. For example, only six plant proteins, consisting of 2S albumin, 7S and 11S globulins, peptidase A1, and a nonspecific lipid transfer protein, were identified in waterlogged grape seeds from medieval York, England, and none from Byzantine Lecce, Italy.[429] Overall, protein preservation of ancient seeds has been found to be relatively low compared to other biomolecular classes, such as carbohydrates and lipids,[429−431] but improved methods of protein recovery and recent successes in identifying plant proteins from residues[99] and stains[432] warrant renewed efforts to analyze ancient plant proteomes from plant macroremains.

3.3. Metaproteomes

Metaproteomes are protein mixtures deriving from more than one proteome. The vast majority of metaproteome research focuses on the study of mixed microbial communities, such as those found in host-associated and environmental micro-biomes,[433] but many of the same methodologies also apply to mixed proteomes of made or manufactured origin, such as food crusts and artist materials, and so these materials will also be considered here. Also considered here are proteins recovered from microbially infected and diseased tissues. Studies of metaproteomes are among the most exciting and fastest-growing applications in paleoproteomics, but they also pose unique challenges for protein identification and authentication.

3.3.1. Microbiomes.
Microbiomes are diverse microbial consortia that form stable complex communities. They may be host-associated, such as the oral and gut microbiomes, or environmental, such as the soil microbiome, and they can vary spatially within a given site or across different conditions. For example, within the oral cavity, the microbiota present on the gums differs from that on the tongue which further differs from that within dental plaque, and even within dental plaque, there are differences between supragingival and subgingival dental plaque.[434,435] Likewise, soil microbiota differ greatly across different environmental conditions, such as deserts, forests, and agricultural fields.[436] In spite of, or perhaps because of, this enormous capacity for diversity and variation, the study of modern and ancient microbiomes is highly valuable.[437]

The first metaproteome of an ancient microbiome was characterized from human dental calculus,[92] a form of calcified dental plaque that forms naturally during life.[438] Numerous microbial proteins were identified, including virulence factors specific to periodontal pathogens, such as TfsA and TfsB from *Tannerella forsythia*, which were independently confirmed using metagenomic techniques. However, despite searching the metaproteomic data against all of UniProt, it was clear that oral microbial proteins were being underidentified due to insufficient representation of oral bacteria in the database. The creation of a custom database based on the translation of genomes in the HOMD[222] produced improved microbial results, and another custom database created by translating metagenomic data from the same samples yielded even more microbial identifications, but the latter approach failed to yield a better understanding of the metaproteome since most of the translated sequences lacked annotation or were categorized as "hypothetical" proteins. Subsequent investigations of dental calculus have confirmed that it preserves an extremely rich and diverse microbial metaproteome,[439,440] but underdeveloped databases (both proteomic and genomic) limit the number of identifications and interpretations that can be made at present.

Beyond microbial proteins, dental calculus also preserves a rich mixed proteome of saliva and gingival crevicular fluid, including proteins involved in host immunological response to dental plaque and tissue destruction due to periodontal disease.[92] Many of the host proteins identified in both modern and ancient dental calculus are expressed by neutrophils,[92] the major cell type involved in periodontal innate immunity.[441,442] Nonimmunity related host proteins in dental calculus include α-amylase, a starch digestive enzyme expressed in saliva.[92]

In addition to microbial and host proteins, dental calculus also variably contains dietary proteins.[149] The first dietary protein to be identified in ancient dental calculus was beta-lactoglobulin, a protein highly specific to milk.[235] Milk proteins were subsequently identified in archaeological dental calculus from Europe,[149,440,443,444] Africa,[223] and Asia,[101,445] contributing to an understanding of how dairying arose and spread in prehistory. Plant proteins have also been identified in dental calculus, and the number and diversity of identified plant

proteins have steadily grown as both methods and instrumentation have improved. The first study to recover plant proteins from dental calculus surveyed the Iron Age to medieval periods in Britain and identified dietary proteins from oats, peas, and cruciferous vegetables only in the youngest samples.[149] Subsequent studies in the Bronze and Iron Age Levant identified not only dietary staples, such as wheat and sesame, but also spices, oilseed, and fruit likely introduced through long-distance trade.[102] Other dietary proteins that have been identified to date in archaeological dental calculus include chicken egg ovalbumin[444] and ruminant hemoglobin.[149]

Although the majority of ancient microbiome research has focused on dental calculus, paleofeces are also now being explored. A recent study of Alaskan dog paleofeces, for example, identified a wide range of host proteins, including proteinases, peptidases, and lipases related to gastrointestinal digestion, as well as dietary proteins deriving from Salmonidae fish that indicate the consumption of fish muscle, guts, and eggs.[446] Few bacterial proteins were identified, as these were intentionally depleted prior to analysis. Cytochemical staining indicates that proteins are present throughout paleofeces,[447] and prior work on paleofeces using immunoassays has suggested that gastrointestinal parasites are also accessible through proteomic techniques.[448−451] Future analyses of paleofeces may yield insights into the structure and function of the ancient gut microbiome, as well as reveal richer information about health and diet.[452,453]

3.3.2. Residues, Crusts, and Food Remains. Ceramic cooking vessels have been shown to preserve dietary lipids and small metabolites (e.g., miliacin, tartaric acid) over long periods, allowing the tracking of prehistoric food practices, such as fish processing,[454−456] the spread of millet,[457] dairying,[458−460] vegetal oil storage,[461] and wine production.[462−464] However, efforts to recover dietary proteins from similar vessels have been met with only limited success to date.[465−467] It appears that proteins generally do not persist in, or cannot be extracted from, pottery.[133,468]

In contrast to the ceramic fabric itself, food crusts provide a much more promising target for molecular analysis,[456,469] and calcified food crusts appear particularly suitable for proteomic analysis. A recent proteomic study of 8,000-year-old calcified deposits (limescale) built up on the interior surfaces of cooking vessels at the Anatolian Neolithic site of Çatalhöyük yielded a remarkably diverse range of food proteins,[99] including milk whey (e.g., β-lactoglobulin, α-lactalbumin), curd (e.g., α, β, and κ caseins), and fat globule-associated proteins (butyrophilin subfamily 1 member A1), as well as meat/blood proteins (hemoglobin) and a wide variety of plant proteins (e.g., hordeins, legumins, serpin-Z4) from cereals and legumes. Dietary proteins were found to be concentrated within the calcified crusts on the interiors of the vessels, and the analysis of noncalcified ceramic fabric from the same vessels produced comparatively few proteins. Importantly, this study showed that ancient cooking vessels were used and reused to cook a wide range of plant and animal foods, with few vessels suggesting specialized use. Although lipid analyses, especially those based on isotopic analysis of $C_{16:0}$ and $C_{18:0}$ fatty acids, may be confounded by such culinary practices, proteomics is ideally suited for sorting out and distinguishing these food mixtures.[99]

Beyond food crusts, foods have occasionally survived relatively intact under exceptional environmental conditions,

such as those found in the Taklamakan desert in western China. There, whole pieces of kefir-like cheese were found adorning mummies associated with the Bronze Age Xiaohe horizon,[107] and dried milk was identified lining the interiors of grass baskets.[108] In addition, sourdough bread made of barley and millet was preserved at the Subeixi cemetery during the subsequent Iron Age period.[95] LC−MS/MS analyses of these remarkably preserved foods yielded proteins not only integral to the food itself but also provided insights into the fermentation of these foods by lactic acid bacteria and yeasts, providing an unprecedented glimpse at prehistoric culinary technologies. Other examples of ancient foods directly analyzed by proteomics include gut content studies of unusually preserved corpses, such the Tyrolean Iceman Ötzi, an Alpine glacier mummy dating to the Chalcolithic,[420] and Tollund Man, an exceptionally well preserved bog body from the Danish Early Iron Age.[470]

3.3.3. Infections and Diseased Tissues. Ancient DNA has contributed to major advances in paleomicrobiology and pathogen genomics,[437,471] and has led to the molecular identification and characterization of more than a dozen infectious pathogens, including *Yersinia pestis*, *Mycobacterium tuberculosis*, *Mycobacterium leprae*, *Helicobactor pylori*, *Treponema pallidum*, *Salmonella enterica*, *Plasmodium falciparum*, hepatitis B virus (HBV), and variola virus.[472] By comparison, paleoproteomic studies of ancient pathogens are in their infancy, but offer the potential to study disease pathophysiology and clinical presentation.

Most proteomics-based studies to date have focused on infectious diseases that produce visible pathologies in the skeleton, particularly tuberculosis. However, initial enthusiasm for the identification of *M. tuberculosis* proteins by PMF[473,474] was later tempered by LC−MS/MS studies that showed a lack of peak specificity and the difficulty of distinguishing *M. tuberculosis* peptides from those of other soil mycobacteria.[475] Other studies tried to achieve higher specificity by applying an immunoassay approach, but lacked controls for environmental mycobacteria.[476] Recently, promising results were obtained for *M. leprae* infection using a combination of paleogenomic and paleoproteomic techniques applied to dental calculus.[477] In this study of a middle-aged adult female with osteological indications of leprosy, a 6.6-fold coverage *M. leprae* genome was reconstructed using ancient DNA techniques, and LC−MS/MS analysis recovered evidence of four mycobacterial proteins. Although none of the mycobacterial peptides were specific to *M. leprae*, taken together the osteological, genomic, and proteomic data provide a convincing portrait of a leprosy infection. Beyond its findings, the study highlights the difficulties of studying infectious diseases using a proteomics approach. Just as the field of ancient DNA only overcame its challenges in paleopathology with the transition to whole genome sequencing,[478] the field of paleoproteomics will also need to achieve greater pathogen proteome coverage before its full potential can be reached.

Beyond pathogen proteins, there has been recent interest in characterizing innate immune proteins within mummified and skeletal remains as a proxy for inflammation and disease.[364,418] While promising, more work is needed to understand the natural levels of immune proteins within healthy tissue, both from modern and ancient contexts, before such findings can be fully interpreted.

3.3.4. Cultural Heritage Materials and Works of Art. The application of mass spectrometry and other methods for
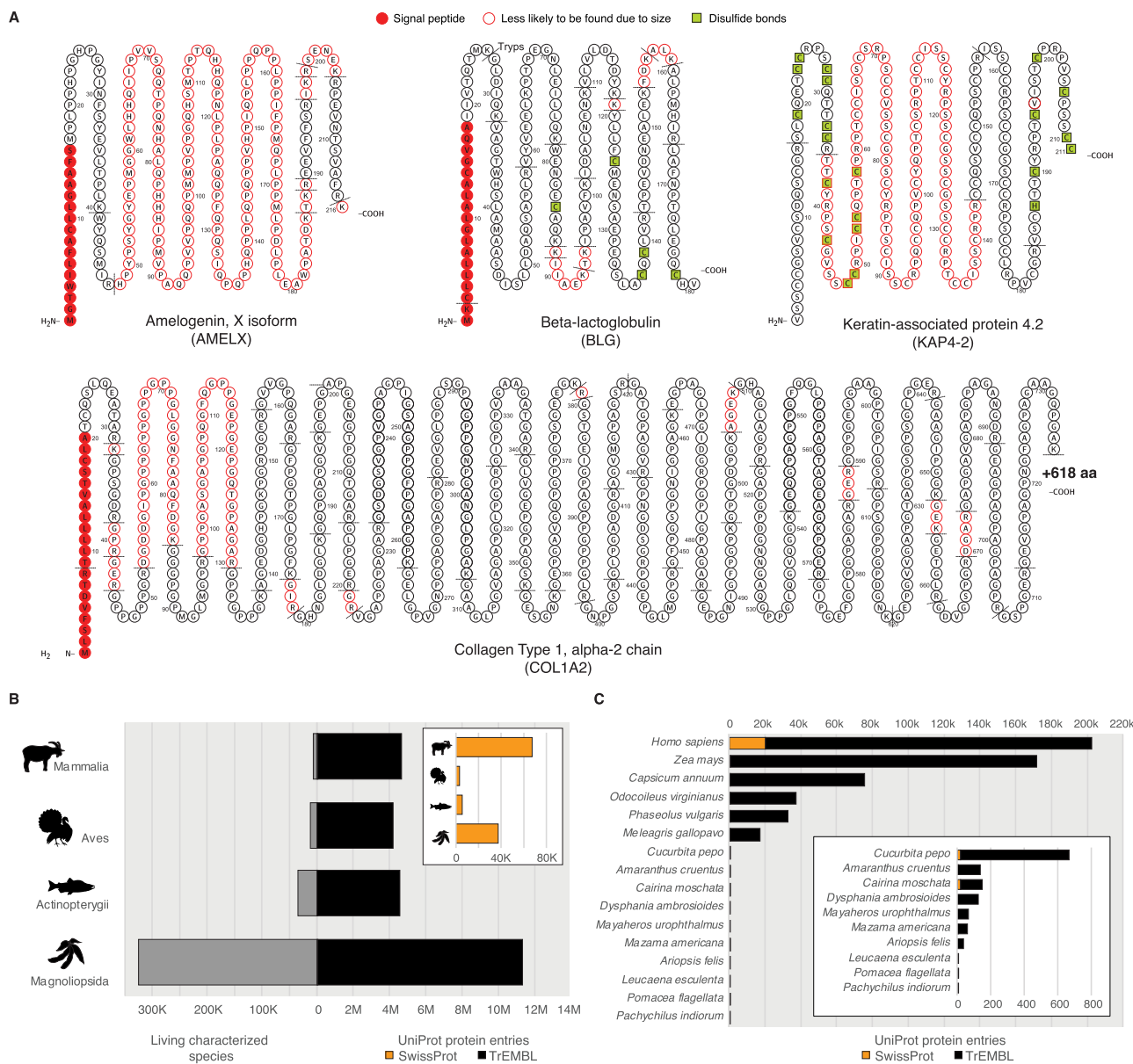
Figure 4. Sources of bias in ancient protein identification. Differences in the density of enzymatic cut sites, number of disulfide bonds, and protein length influence protein detectability (A). The representation of proteins across taxa is highly uneven in major databases, such as UniProtKB (B,C). (A) Comparison of four *Ovis aries* (sheep) proteins of archaeological interest: AMELX, used for determining sex; BLG, a milk protein; KAP4-2, a protein component of wool; COL1α1, a bone protein used for taxonomic identification. Selected properties influencing protein detectability using LC−MS/MS techniques include predicted peptide size following trypsin digestion (cut sites are shown as dashed lines) and location of disulfide linkages (green boxes). Peptide lengths of 6−30 amino acids are most suitable for detection,[165] and those outside this range are marked as less likely to be found due to size (red border). Note that endogenous proteolysis of amelogenin during dental development causes additional cleavages that are not shown.[603] Only half (736 amino acids) of the collagen protein is shown for space reasons. (B) Comparison of the number of characterized species to the number of protein entries in UniProtKB (SwissProt and TrEMBL) for the major taxonomic classes Magnoliopsida (flowering plants), Mammalia (mammals), Aves (birds), and Actinopterygii (ray-finned fishes); inset shows enhanced view of the number of reviewed (SwissProt) protein entries. Numbers of characterized species were obtained from refs 604−607. (C) Reviewed and unreviewed protein entries available in UniProtKB for humans and common plants and animals consumed in ancient Mesoamerican diets; inset shows enhanced view of taxa with <1000 protein entries.

inferring proteinaceous components of artwork and other cultural heritage objects has been comprehensively reviewed elsewhere.[479−483] Both MALDI-TOF and LC−MS/MS can be used to analyze artwork and cultural heritage materials, and their uses are described briefly below.

In ancient and historic times, proteinaceous materials (e.g., milk, eggs, blood, and gelatin from skin or bone) were widely used as binders for pigments in works of art and for building

materials, such as mortars. Understanding the composition of these materials allows insight into past practices and provides information that will inform curation choices. Proteomic analysis of paintings has allowed identification of many widely used binder proteins, including caseins and beta lactoglobulin from milk, collagens from gelatin, vitellogenins, apolipoproteins, and low-density lipoprotein receptor from egg yolk, and ovalbumin, ovotransferrin, and lysozyme from egg white. While

some studies have found that binders from a single source were used,[89,98,157,484−488] many binders are composed of mixed sources, such as milk from at least two different species or various combinations of milk, eggs, and gelatin.[97,199,489−497] Proteins introduced unintentionally during the creation of binders, such as muscle or blood proteins in collagen-based glues, might be able to provide insight into gelatin production or the addition of other animal products into the binders.[489] Beyond paints, binder proteins were also added to construction materials, such as mortars. Blood and milk are frequently recorded as mortar additives, and they have been detected in ancient mortars using a variety of techniques[498−500]

There are many historically recorded recipes for binders, and proteomics can illuminate the use of different recipes as they correlate with pigment color, type of canvas or statue, properties and availability of the binder, and cultural and personal choice.[97] Currently this is limited to some extent by detection bias and the limited number of samples analyzed. Because caseins and some egg proteins are not detected well with trypsin alone, multienzyme digests using both trypsin and chymotrypsin can improve detection of these binding materials.[169] In cases where the proportion or preservation of binder proteins is very low, targeting techniques such as MRM can help improve the detection of binder peptides, and this has proven particularly useful for identifying egg proteins in cases where their presence was suspected but not previously detected.[199] The recent development of low cost, minimally destructive sampling techniques that can be conducted without specialist training[154,157,169,490,501−503] promises to make proteomic studies of artwork more feasible, and further development of PMF using MALDI-TOF will facilitate larger scale data collection.

In addition to paint and mortar binders, a number of other historical and archaeological cultural heritage items have also been analyzed using proteomics techniques. Collagenous glues made from both mammal and fish gelatin have been identified in historical and archaeological samples.[201,504,505] Many other varying items have been analyzed including cosmetic sticks,[506] metal coated gut threads,[507] organic coatings on skulls and artifacts,[508] a birth girdle,[432] photographs,[503] and collections of museum items.[254] Proteomic analysis of these items can assist in better understanding the knowledge, processes, and choices used to create material culture.

Proteomics is also proving useful for understanding past conservation practices and guiding future efforts. For example, a recent study of historically conserved calfskin revealed cross-linked calf-rabbit collagen peptides, indicating that form-aldehyde and rabbit glue had been used to conserve the piece.[509] In addition, identification of mixed and even mislabeled glues as part of conservation practices is just beginning to be explored.[490] Proteomic techniques can also help identify fungi and bacteria living on the surface of cultural heritage objects, such as parchment,[510] and in future it may be possible to use this information to identify objects at risk and to enhance methods of conservation.

4. CURRENT CHALLENGES

Paleoproteomics is revolutionizing the way we study the past and providing unprecedented insights into evolution, phylogeny, human economies, cuisine, art, and other forms of material culture. However, current inefficiencies in protein recovery and measurement, limits of instrumentation, and insufficient databases and computational power hamper our ability to fully access ancient proteomes. Combined, these problems can result in the failure to identify key proteins that are present in a sample, create biases of detection, or limit the ability to quantitatively analyze archaeological and historical proteomes. Here we discuss major challenges still to be overcome with respect to the detection, identification, and authentication of ancient peptides and proteins.

4.1. Protein Detection

In order for proteins to be detected they must become incorporated into the archaeological record, survive over time, denature, solubilize, and be digested during the extraction process, ionize in the mass spectrometer, and fall within the instrument's dynamic range of detection. These are hard limits for mass spectrometry, and any peptides that remain undetected are unrecoverable. Biological and diagenetic variation can lead to proteins and peptides being unequally incorporated, degraded, extracted, digested, and ionized in ways that are influenced by the protein's amino acid sequence, hydrophobicity, protein structure, chemical modifications, interactions with a mineral matrix, and choice of extraction method.[9,140,366,511,512] Instrument design and performance also limits the detection and reproducibility of findings.[92,188,191,513,514] Some of these factors can be anticipated, while others not.

Biases related to protein size and sequence are among the most predictable. For example, some proteins (e.g., COL1) are extremely large, and thus produce many peptides, making their detection more likely than smaller proteins, even if both proteins are present in equal quantities (Figure 4A). Other proteins, such keratins, have a large number of cysteine disulfide linkages that can be difficult to fully break during extraction, resulting in underdetection (Figure 4A). Following digestion, some peptides can be predicted to have low recoverability because they are either too small or too large for efficient ionization and measurement by a given instrument (Figure 4A). In addition, diagenetic changes can also alter protease cut sites, changing detectability. For example, the positively charged side chains of Arg and Lys are prone to deamination and glycation, which contributes to missed cleavages and reduces the efficiency of trypsin in older samples. *In vivo* autodigestion (e.g., enamel proteins) and diagenetic backbone cleavage also alter expected peptide profiles.

Peptides from a given protein that are able to be consistently detected are known as proteotypic peptides; however, despite great effort, attempts to predict proteotypic peptides have been met with only limited success.[515−517] The inability to adequately correct for detection biases across the proteome represents a major challenge for protein quantitation[518] and is even more challenging for ancient samples and discovery-based applications.

One of the most vexing problems in shotgun proteomics is its limited reproducibility, especially for low abundance proteins, which are often the main proteins of interest in ancient protein studies. Two factors combine to contribute to this problem. First, gene expression, and hence protein abundance, typically varies by more than 6 orders of magnitude within a given tissue,[358,359,519] and even up to 12 orders of magnitude within biofluids.[520,521] This is starkly different from DNA, where most genes are present in a single copy and even multicopy genes rarely differ in abundance by more than 1 order of magnitude.[522] Thus, the dynamic range of a proteome

is enormous. Second, current tandem mass spectrometers are not powerful enough to selectively fragment and analyze all precursor ions generated from a given sample,[514] and thus only a fraction of the complex peptide mixture is accessible by conventional shotgun proteomics.[523] For a given mass-to-charge (m/z) window, current instruments either select the most abundant ion for fragmentation and MS2 (DDA) or they select all ions, producing a mixed MS2 (DIA).[204] Neither are satisfying solutions for ancient proteins. DDA leads to poorly reproducible ion selection and underselection of low abundance peptides, while DIA produces MS2s that are difficult to deconvolute and still suffer from low signal-to-noise ratios for low abundance peptides. Targeting methods, such as MRM or PRM, only partially address the problem and are wasteful of sample, as they only measure preselected peptides,[524] which can be difficult to know in advance for discovery-based studies or when diagenetic modifications of target proteins are unknown. Better chromatography, ion mobility, faster, more powerful instruments, and computational improvements may soon ease some of these problems,[191,525−529] but at present ancient protein studies have yet to achieve the high degree of reproducibility that paleogenomics enjoys.

4.2. Protein Identification

Protein identification, especially from complex mixtures of unknown composition, remains an ongoing challenge. Sequence-based identification remains the gold standard in paleoproteomics, but such identification is strongly influenced by the choice of database and anticipated modifications, meaning that only peptides that are looked for can be identified. In the light of these limitations, some analysis strategies include error-tolerant or *de novo* sequencing approaches in order to increase the number of identified MS2 spectra beyond that of the database. However, as the sample search space increases through the addition of sequences or permitted modifications, computational demands can quickly exceed current feasibility. Fortunately, unlike detection limits, identification limitations are soft in that data can be later reanalyzed with updated software, algorithms, databases, computing infrastructure, and chemical modification parameters to improve identifications.

Databases are currently a major limiting factor in ancient protein identification. Although the proteomes of humans[359] and some model organisms[519,530,531] are now well curated and annotated, the proteomes of many taxa of archaeological interest, from molluscs to microbiomes, remain insufficiently or poorly characterized. Most databases suffer from inclusion bias, with model organisms and economically important taxa being vastly overrepresented compared to other species. Domesticated taurine cattle (*Bos taurus*), for example, has a published genome and a well-annotated proteome, with 140,740 entries in UniProtKB and 9,936,498 in NCBI GenBank. In contrast, only limited genetic sequence data are available for the banteng (*Bos javanicus*), a related species of cattle from Southeast Asia for which there are fewer than 400 protein sequences in NCBI GenBank and UniProtKB. Even in cases where a full genome for a species is available, there can be significant delays between when genomic data is submitted to a database, such as NCBI GenBank, and when its annotated proteome becomes available across linked platforms, such as UniProtKB. For example, at the time of writing, UniProtKB contained only 97 reviewed and 2,985 unreviewed proteins for

domesticated water buffalo (*Bubalus bubalis*), but its entire genome was available through NCBI GenBank, including 64,378 annotated proteins. In addition, although UniProtKB aggregates data from major genetic databases, it does not incorporate all annotated genomic data, and some curated protein databases continue to be independent, such as the HOMD, WormBase, ParaSite, and VectorBase. Thus, protein representation even within large protein databases, such as UniProtKB, is highly variable across major taxonomic groups, especially when accounting for the number of known species per group (Figure 4B), and careful consideration must be given to database selection and database biases during study design.

Several commonly observed data artifacts in paleoproteomics studies are database-dependent and can be avoided by accounting for the known biases of specific databases and search algorithms. For example, where there is a substantial imbalance in the amount of protein sequence coverage for related taxa, database searching may incorrectly identify the protein as originating from the wrong species due to the higher number of conserved peptide matches to the better covered species, which may have more complete sequences as well as more isoforms. This is commonly observed in studies of domesticated bovids, which are frequently targeted for taxonomic discrimination using proteomic methods. Taurine cattle have roughly 1.5 times the amount of protein entries in UniProtKB than sheep (91,483) and goat (90,609) and the imbalance is even greater in NCBI GenBank, with taurine cattle having 20 times the amount of protein entries as sheep (336,178) and 40 times the amount of protein entries than goat (156,078). Proteomic analyses of these and other less well studied bovids generally return at least some protein taxonomic assignments to taurine cattle, the latter being an artifact of the more extensive representation of taurine cattle proteins in the database. While some taurine cattle protein entries in these databases are redundant, others contain important allelic sequence variants, and sequences generated from transcriptome data additionally include isoforms deriving from alternative splicing that are highly relevant for proteomics studies. When attempting to use proteomics to distinguish between closely related species, not only is the coverage of each species in the protein database important, so is the difference between the number of protein entries present.

Another common artifact occurs when databases contain entries for a protein with differing degrees of completeness, and entries containing only the mature protein sequence are preferentially (and sometimes incorrectly) identified over entries that also contain the signal peptides and other regions that are removed during mature protein formation. This has been observed in studies of mammalian collagen when search algorithms return spurious high-scoring assignments to extinct species, such as the giant South American ungulate *Toxodon*. In the case of *Toxodon*, the fossil COL1A1 protein entry in UniProtKB consists of only the helical region of the protein (UniProtKB C0HJP7, ~1000 amino acids), whereas COL1A1 entries for extant taxa also include the flanking signal peptide, propeptide, and telopeptide regions, making it ~500 amino acids longer. Signal and propeptides are removed during mature protein formation and are therefore never recovered from archaeological samples, while telopeptides are only rarely recovered. Because collagen is highly conserved across mammals, many recovered peptides will be shared among all ungulates, but these conserved peptides will have a higher percentage of protein coverage to *Toxodon* COL1A1 compared

to extant taxa, making *Toxodon* a higher ranked match. Such artifacts can be easily identified and corrected by manually checking the sequence alignments and verifying the taxonomic specificity of the matches. Thus, sequence and taxonomic validation are important steps of data analysis in exploratory paleoproteomics studies.

Beyond issues of database completeness, databases also contain entries with varying levels of quality, annotation, and associated metadata. Within UniProtKB, the vast majority of sequences are unreviewed (TrEMBL), meaning that they are generated through an automated annotation process, while proteins whose annotations are manually reviewed (SwissProt) make up <0.5% of currently available protein data for flowering plants, mammals, birds, and fish (Figure 4B). While automatically annotated sequences in TrEMBL are largely accurate, they often require further review of the source entry and comparison with the same gene in other species. Sometimes it can be helpful to search against translated nucleotide databases (tblastn) in order to assess the range of species in which similar sequences have been identified. Many unreviewed entries also lack sufficient metadata for downstream analysis and may be simply annotated as a hypothetical or uncharacterized protein.

Collectively, these problems can be particularly acute for paleodietary studies in which ancient groups consumed a wide variety of foods that may or may not have adequate database representation. Consider, for example, the components of a hypothetical ancient Mesoamerican diet shown in Figure 4C, which includes common foods consumed in Central Mexico and the Maya lowlands. The human proteome is also included for comparison. At the time of writing, the number of total protein entries in UniProtKB for maize (*Z. mays*, 171,947) vastly exceeded that of other staple grains, such as amaranth (*A. cruentus*, 138), the database representation of white-tailed deer (*O. virginianus*, 37,513) and dwarfs that of red brocket deer (*M. americana*, 59). Protein entries for turkey (*M. gallopavo*, 17,051) greatly exceed those of muscovy duck (*C. moschata*, 146), as do entries for common beans (*P. vulgaris*, 32,845) compared to other legumes (*L. esculenta*, 9). Other food items, such as common vegetables (*C. pepo*, 667; *D. ambrosioides*, 124), local fish (*A. felis*, 37; *M. urophthalmus*, 65), and edible snails (*L. esculenta*, 9; *P. flagellata*, 4; *P. indiorum*, 1) have very few protein entries, effectively making them invisible in proteomic studies of ancient diets. Using an alternative database, such as NCBI RefSeq or GenBank, substantially improves the protein representation for some foods (e.g., *C. pepo*, 43,466) but not others (e.g., D. ambrosioides, 250; A. felis, 61; and *P. indiorum*, 1). Thus, while changing or combining databases can improve the identification of some genetically well-studied taxa, little can be done to improve the visibility of many other ancient Mesoamerican foods until more genetic data is available. Paleoproteomic characterization of ancient diets is therefore largely opportunistic. As such, while the identification of a dietary protein can be taken as positive evidence of its presence, the failure to identify a given food cannot be taken as evidence of its absence.

Beyond using existing protein databases, mining sequences from genetic and transcriptomic databases (proteogenomics) that are not automatically annotated by UniProtKB to create custom databases can increase identifications, as can *de novo* sequencing of proteins or peptides of interest. However, these options can be time-consuming and low-throughput if the proteins of interest are not already known or if the taxa themselves are understudied, which limits their feasibility in discovery-based applications. Massive gains in annotated genomic sequence data are needed to overcome this problem, and large-scale international efforts to dramatically increase the cataloging and characterization of eukaryotic biodiversity, such as the Earth BioGenome Project,[532] the Vertebrate Genomes Project,[533,534] and the Darwin Tree of Life Project,[535] offer great promise for improving the detection of dietary proteins in future paleodietary studies.

For now, however, even taxonomic identifications based on extremely well characterized proteins, such as COL1, face challenges. For example, there are currently no curated databases for COL1 PMF markers, which poses a large barrier to the exponentially growing ZooMS community. Beyond the taxonomic biases of proteomic databases, such as UniProtKB, and especially the underrepresentation of fish and birds, COL1 genes are often incorrectly translated from genetic data or contain incorrect annotations. In the absence of sufficient databases, multivariate analysis and other statistical tools have been successful at clustering COL1 and keratin peptide groups for taxonomic identification of sample types such as bone and hair.[315,536] Moreover, in the case of COL1 the charge pairing in the molecule required for fibril formation results in highly conserved tryptic cleavage sites, meaning that it is possible to compute all potential variants for every arginine- and lysine-terminated peptide, generating theoretically huge databases of 10^{222} collagen sequences,[20] and thereby potentially, albeit impractically, allowing the *de novo* identification of COL1 peptides with high fidelity from MS2 spectra.

As the field of paleoproteomics progresses, more consistent reporting of search settings and the percentage of identified spectra will help to highlight the extent of the problem of identification, as well as allow metaanalysis of the success of different search strategies for different types of samples. With the exponential growth of genomes, transcriptomes, and proteomes, increasing characterization of modifications, and development of new PMF markers, reanalysis of previous samples will likely result in an increasing amount of detected proteins and/or taxonomic resolution. However, this requires publication of raw data to facilitate reanalysis, as is already recommended for ancient LC−MS/MS data[4] and facilitated by the ProteomeXchange consortium[537] via repositories such as PRIDE.[538] As no global repository exists for MALDI-TOF data, it is currently less common for PMF raw data to be published, although growing numbers of data sets are found on the open-access general purpose repository Zenodo (https://zenodo.org/search?q=zooms), which was developed under the European OpenAIRE program and is run by CERN.

4.3. Protein Authentication

One challenge that particularly affects paleoproteomics is the difficulty of distinguishing authentically ancient peptides, proteins, and proteomes from environmental and modern contamination.

4.3.1. Sources of Contamination. Potential sources of contamination are myriad, but for modern proteomics they are mostly lab-derived and easily mitigated. Contamination poses a much more challenging problem for paleoproteomics, where the proteins of interest are degraded, in low abundance, and frequently unknown. Sources of contamination from the laboratory are well described[135] and can be identified and controlled through the analysis of extraction and instrument blanks and the application of best practices.[4]
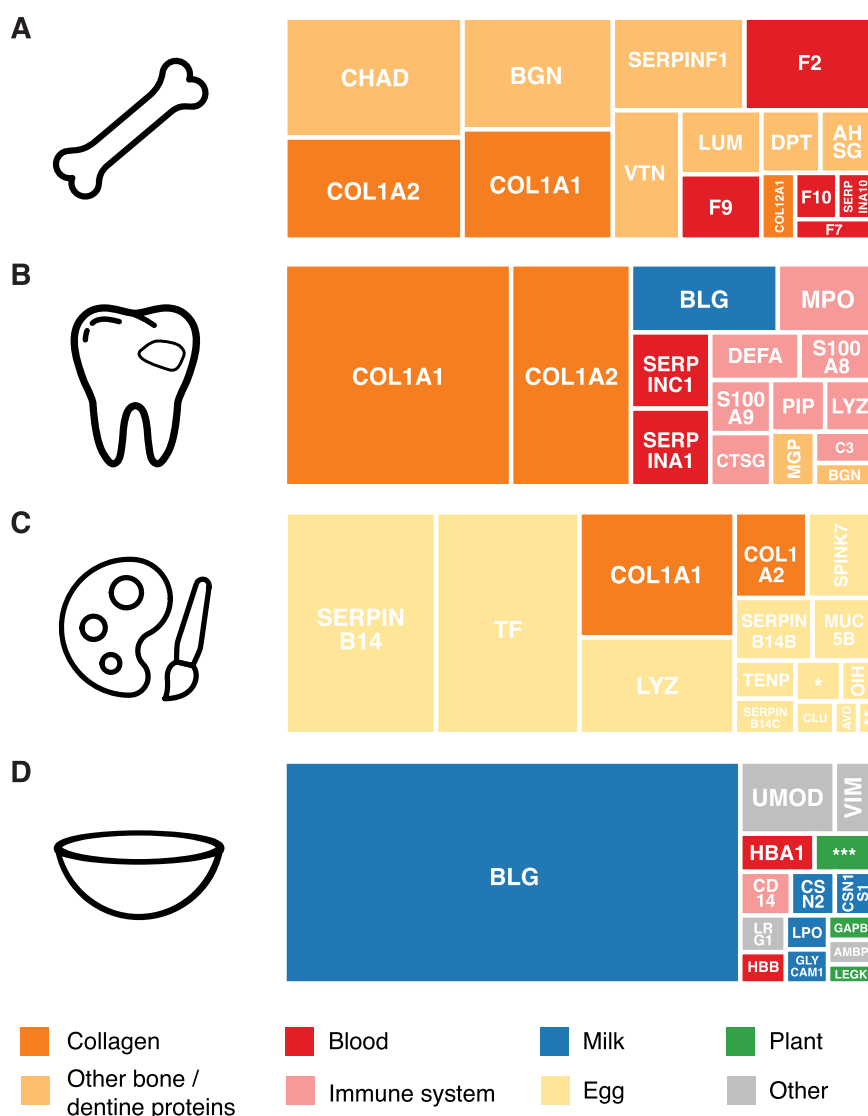
Figure 5. Representative examples of ancient proteomes. Well-preserved ancient proteomes contain distinctive groups of proteins that reflect the protein composition of the original tissue or material, such as human bone[364] (A), human dental calculus[445] (B), artist materials[137] (C), and pottery crusts[99] (D). As such, the composition of an ancient proteome can aid in its authentication. Data were searched against the SwissProt database using Mascot using the parameters described in ref 102. Protein identifications were established at <5.0% protein FDR and <1.0% peptide FDR in Scaffold v.5 (Proteome Software), and proteins with a minimum of 97% protein identification probability and at least two unique peptides were accepted. The top 15 proteins (by number of PSMs) per sample source were visualized as a treemap and labeled by their corresponding gene name; trypsin, keratins, serum albumin, and microbial proteins were excluded from the analysis. *Ovostatin; **riboflavin-binding protein; ***B3-hordein.

Less controllable are environmental contaminants that are introduced over time by the burial or depositional context. Environmental contamination can be accounted for to some degree by collecting and analyzing control samples from other associated remains or nearby sediment or soil, although this is not always feasible. The inclusion of controls in the experimental design is particularly important for paleoproteomic studies on: (1) open systems, such as the identification of a nursing female dog from milk proteins found in rib bones;[539] (2) sediments or soils, such as the identification of silk from textile imprints in the soil;[332] and (3) infectious pathogens, especially taxa that have close environmental relatives, such as *Mycobacterium tuberculosis*.[437,475]

Even more variable and often unpredictable are contaminants introduced during excavation, handling, storage, and conservation. This is a particular problem when analyzing

museum collections where handling and curation history are not well documented. Contamination can occur during handling when collections are used for teaching or placed on display, and contaminants can also be unintentionally introduced through local storage conditions and may include proteins from bacteria, fungi, and rodent or insect pests, as well as treatments used in their control.[540] Object conservation is also a major source of contamination. Although today museums and conservators have synthetic options for consolidants, adhesives, and preservatives, natural products were almost exclusively used in the past, and some continue to be used as they provide certain benefits over synthetic materials.[541,542] Those most relevant for paleoproteomics research are collagen based glues,[543] fibroin used to repair silk,[544] and egg and milk based glazes and treatments, such as the application of egg whites and phytic acid as a flame

retardant.[545] While the impact of conservation efforts on ancient DNA, stable isotope, and radiocarbon analyses has been extensively investigated,[546−548] their effect on paleoproteomics has been less systematically studied. Best practices involve avoiding samples where treatment or extensive handling has taken place.

4.3.2. Methods of Authentication. There are two main approaches to authenticating ancient proteins. The first involves the identification of damage patterns characteristic of ancient proteins. The second examines the broader context of the proteins, including its associated metaproteome and other corroborating lines of evidence.

Among ancient biomolecules, DNA undergoes regular and predictable forms of degradation that produce characteristic types of ancient DNA damage,[549,550] the most important of which are DNA fragmentation and cytosine deamination.[551] Their mechanisms of formation are well understood, and they are so consistent that they can be used to authenticate ancient DNA[552] and even distinguish ancient and modern DNA sequences in heavily contaminated samples.[553] Proteins also undergo processes of degradation, and methods have been proposed to authenticate ancient proteins or proteomes in cases where sufficient numbers of identified MS2s are available for statistical analysis (e.g., ref 554). The biochemical and structural complexity of proteins, however, makes damage-related authentication of ancient proteins considerably more complicated than for ancient DNA.

For very ancient samples, hydrolysis is expected to cause protein fragmentation, resulting in a bias toward shorter enzymatic peptides, and an increased number of nontryptic cleavage sites (reviewed in[127]). Diagenetic amino acid modifications have also been proposed as markers of authenticity, but due to the increased biological complexity of proteins, other factors besides age can strongly influence the production or prevention modifications.[9,555] In particular, deamidation of asparagine and glutamine have been proposed as markers of authentication in both PMF and LC−MS/MS data.[275,365,554,556−558] Unlike deamination in aDNA, however, deamidation also occurs *in vivo*[559] and is strongly impacted by both local depositional chemistry[560] and choice of extraction method.[140,561] Therefore, its use as a reliable age-related indicator has been questioned.[555,560,562,563]

In MALDI-TOF applications, COL1 deamidation has been used as a relative age marker and proposed as a criterion to identify intrusive samples. In practice, however, this requires very large data sets, and even when data for large numbers of samples (>2,000) are available, the accuracy of seriating samples by relative age class is less than 50%.[560] For LC−MS/MS studies, deamidation has mostly been analyzed semi-quantitatively at the metaproteome level (e.g., refs 14, 92, 197) in order to show that deamidation is a top modification in the data set. Some studies have attempted to authenticate specific proteins using deamidation patterns,[101,439] but this requires a relatively large data set to perform statistical analysis, which is not always possible,[223] and even in cases where enough data is available, the results are not straightforward to interpret.[562] Nevertheless, proteins of ancient origin should exhibit evidence of diagenesis. Caution should be exercised if reports contain evidence of unexpected chemical behavior such as the lack of any degradation of proteins over extended periods of archaeological or geological time[564] or if unexpected proteins are detected that are commonly used in the same facility.[565] Future improvements in understanding and modeling protein

diagenesis at specific amino acid sites may one day enable protein damage to be used more reliably and quantitatively as an age indicator. For now, however, it is most effectively used as a qualitative indicator of age in combination with other authentication approaches.

Contextual analysis focusing on the composition of the metaproteome as a whole, together with other corroborating lines of evidence, is currently the most robust authentication approach for ancient proteins. One important feature of proteins is that their expression is tissue specific, and thus the composition of the ancient metaproteome itself can be used to aid in its authentication. Different tissues and substrates, such as bone, dental calculus, artist materials, and pottery food crusts, are expected to each have a different and distinctive protein composition, and this is reflected empirically in ancient samples (Figure 5). Clear patterns are readily apparent. Bone is dominated by collagens and proteins involved in fibrilar organization and mineralization, as well as blood clotting factors and plasma proteins.[364] Dental calculus is enriched in collagen and proteins associated with the innate immune system (especially neutrophils), and also contains salivary amylase and occasional dietary proteins.[92,445] Artist materials, such as paints, tend to be dominated by a single protein source, such as egg,[137] and pottery food crusts are highly diverse but contain elevated levels of animal and plant proteins associated with foods, in this case foods independently known to have been consumed at the site.[99] For sample types that naturally contain a microbial component, such as microbiomes, the bacterial protein and taxonomic composition can also serve as an authentication aid,[92,440,566] and for sample types that do not contain a microbial component during life or use, the relative proportions of microbial proteins to presumed endogenous proteins can also serve a relative preservation indicator.

As more modern and ancient metaproteomes are characterized and their data made public, the characteristic features and composition of well-preserved metaproteomes can be defined and used as benchmarks for paleoproteomic authentication. Such data, in combination with independent corroborating lines of evidence, such as paleogenomic data, paleoethnobotanical evidence, or zooarchaeological findings, can be used to assist in establishing the plausibility of findings, particularly those that are extraordinary or unanticipated.

5. FUTURE DIRECTIONS: TAKING ON THE DARK PROTEOME

Despite significant technological advances, much of the ancient proteome still remains "dark" as seen in the consistently low percentage of identified MS2 spectra. This darkness, however, also represents an opportunity, as it means that there are still ancient proteins and proteomes to be discovered if only the right tools can be developed to access them. On the basis of current knowledge, the dark proteome can be divided into three types: (1) "structurally dark" regions of proteins[567] that are intrinsically disordered, never resolved, or (in the case of bottom up proteomics) lacking suitable cleavage sites; (2) "overlooked" peptides, peptides that have been modified to such a degree that their spectra are not recognizable by conventional MS2 analysis; and (3) "unidentifiable" protein fragments that have cross-linked or condensed into novel chimeric structures that have not been previously studied. Below we discuss new technologies that may assist with understanding and accessing the dark proteome, as well as

consider the potential of emerging nonmass spectrometry-based sequencing approaches in paleoproteomics.

5.1. Emerging Technologies in Mass Spectrometry

Many of the unmatched queries found in ancient protein data sets are presumably the result of poor ion selection, fragmentation, and/or detection. A range of new instrumentation options are being developed that include faster data acquisition, higher sensitivity, and increased ion transfer efficiency. However, mass spectrometers are now reaching the point where chromatography will need to be improved in order for these instrument improvements to actually provide better data,[191] and yet paleoproteomic samples often suffer from poor chromatography due to unknown coextracted compounds or other factors. Alternatively, chromatography may be augmented or even eventually replaced by enhanced gas phase separation and fractionation, such as ion mobility, as is currently being implemented in Direct Infusion Shotgun Proteome Analysis (DISPA).[568] Future research development needs to focus on specific optimization for paleoproteomic samples, with the aim of improving precursor separation, reducing peak width, and generating better MS2 spectra. This will ultimately improve peptide identification and even enable *de novo* sequencing to resolve some of the overlooked peptides. This can be combined with improvements in databases, characterization of proteins, structural determination (e.g., AlphaFold[569]), and digestion[168] to aid in the recovery of structurally dark sequences.

Although chromatography improvements and database expansions represent low hanging fruit, identifying overlooked proteins and novel chimeric structures will be much more difficult, and it will require a better understanding of the diagenetic processes that facilitate preservation but inhibit recovery and challenge identification. In addition to side chain modification, backbone hydrolysis, and cyclization, peptides can undergo condensation with other molecules in the archaeological environment, including other proteins, lipids, carbohydrates, nucleic acids, metabolites, and inorganic compounds. Some of these reaction types are already targets for methodological improvement due to their presence *in vivo* (e.g., glycoproteins and some types of protein−protein and protein-nucleotide cross-linking).[509] However, many are likely novel degradation products or uncharacterized interactions that involve hydrolysis and reformation of the peptide backbone to form truly chimeric proteins.

One approach to this problem is to use experimental time series data to attempt to recreate the process of diagenesis. A time series has the advantage that even if compounds cannot be directly identified, their rise and fall highlights their place in a complex diagenetic pathway. Thankfully, the technologies to analyze these are routine rather than emerging. Raman spectroscopy can give insights into chemical character, amino acid analysis can detect the building blocks, and FT-ICR MS can establish the atomic composition of the masses. Similar approaches are used to understand the humification of soil organic matter[570,571] and the formation of melanoidins, the condensation products of proteins and carbohydrates. Greater structural insights into these diagenetic products will likely reveal a multiplicity of products each present in vanishingly small concentrations much like the hydrocarbon "hump" (unresolved complex mixtures of diagenetic components found in crude oils[572]) or the myriad of compounds identified in humic extracts.

An alternative approach is to visualize the interactions between the different combinations of peptides, proteins, and mineral surfaces that play a role in protecting sequences over deep time, but also likely inhibit recovery. Atomic force microscopy and high resolution 2D electron microscopy and 3D tomography imaging have already begun to inform the intimate association between, and therefore survival mechanisms of, collagen and mineral in bone.[573] Continued investigation into bone, as well as other biominerals and degradation products that produce mineral surfaces, will allow for both a better understanding of protein survival and, more importantly, modifications unique to such protein mineral interactions. Building a greater understanding of these modifications into search strategies will hopefully uncover more peptides, while pattern based discrimination analyses (e.g., machine learning) will assist in identifying groups of new products forming over time.

Spatial subcellular proteomics is still in its infancy,[574] but will likely make considerable advances over the next decade.[575] It combines super-resolution microscopy and top-down proteomic analysis to enable *in situ* mass spectrometry imaging (MSI). Spatial resolution will enhance our understanding of layered and otherwise structurally organized proteins such as cross sections of artwork or incremental precipitation of biologically induced or mediated mineral formation. The best known and highest resolution method is MALDI MSI, which is already providing detailed molecular images of tissues, but it is limited by its need for surface matrix deposition in many cases. Although their resolution is not as good as MALDI MSI, instruments which do not require matrix deposition and may be better suited to the size of paleoproteomic samples are being explored such as *in situ* DESI, nano-DESI, and LESA MSI.[194,195,576] Although current super-resolution microscopy approaches rely on fluorescence based methods[577] which are challenging for ancient samples, this is a growing field of research with new approaches being proposed and developed (e.g., Akoya Bioscience's CODEX, Resolve's Molecular Cartography combined with Zeiss Microscopes) which push the limits of imaging. Even if such imaging approaches are not suited for ancient samples, they can be used to investigate experimental samples to better characterize "overlooked" and "unidentifiable" parts of the dark proteome.

5.2. Beyond Mass Spectrometry?

For the past two decades, mass spectrometry has been the workhorse of proteomics and protein-based studies; nevertheless, mass spectrometry is also limited in that "instead of truly sequencing, it classifies a protein and typically requires about a billion copies of a protein to do it".[67] A variety of other non-MS based technologies are currently emerging that offer the promise of both higher sensitivity and greater scale, such as massively parallel sequencing by nanopore, fluorosequencing, and image-based Edman.[67,68,578] At present, these techniques are not ideally suited to paleoproteomics due to the anticipated and unexpected modifications present in ancient samples,[579] as well as the autofluorescence caused by diagenesis,[580] but as these technologies continue to be optimized, their use may come within reach of paleoproteomics. In addition, the increasing focus on MS-based spatial and single cell proteomics mean that these technologies are far from the only options that will become available over the next decade as alternatives to conventional mass spectrometry.

6. CONCLUSION

The mass spectrometry revolution in paleoproteomics has contributed to unprecedented findings over the past two decades, from the discovery of new archaic human fossils[184,291,292] to the detailed characterization of Neolithic and Bronze Age cuisines.[95,99,101,102,107,108,223,445] Ancient proteins have been particularly valuable for revealing ephemeral aspects of the past, such as the care and repair of art and museum objects,[509] the construction methods of books,[156,295] and even the difficulties of labor and childbirth.[432] While much of the ancient protein research over the past two decades has been exploratory and opportunistic in nature, the data produced and the knowledge gained will allow for larger scale and more directed questions to be explored in the coming decades. With this expansion, paleoproteomics is poised to provide insights from the past that can inform on matters ranging from ecological conservation to human health and wellbeing. Exciting new instruments and capabilities are on the horizon that promise to push the limits of our detection even farther, but only a better fundamental understanding of how proteins both degrade and persist will allow us to access the dark proteome that predominates in most archaeological samples. The past 20 years have been characterized by unprecedented gains in the ability to detect and identify ancient proteins, proteomes, and metaproteomes through advances in mass spectrometry. The next 20 years will surely hold many surprises as we begin to apply this analytical power to answer long-standing questions about the past and innovate new solutions to old problems.

ASSOCIATED CONTENT

Special Issue Paper

This paper is an additional review for *Chem. Rev.* **2021**, volume *121*, issue 19, "Frontiers of Analytical Science".

AUTHOR INFORMATION

Corresponding Author

Christina Warinner — *Department of Anthropology, Harvard University, Cambridge, Massachusetts 02138, United States; Department of Archaeogenetics, Max Planck Institute for Evolutionary Anthropology, Leipzig 04103, Germany;* orcid.org/0000-0002-4528-5877; Email: warinner@fas.harvard.edu

Authors

Kristine Korzow Richter — *Department of Anthropology, Harvard University, Cambridge, Massachusetts 02138, United States*

Matthew J. Collins — *Department of Archaeology, Cambridge University, Cambridge CB2 3DZ, United Kingdom; Section for Evolutionary Genomics, Globe Institute, University of Copenhagen, Copenhagen 1350, Denmark;* orcid.org/0000-0003-4226-5501

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.chemrev.1c00703

Notes

The authors declare no competing financial interest.

Biographies

Christina Warinner is Associate Professor of Anthropology at Harvard University and Group Leader of Microbiome Sciences at the Max Planck Institute for Evolutionary Anthropology. She earned her Ph.D. from Harvard University in 2010 and received postdoctoral training in genomics and proteomics at the University of Zurich (2010-2012) and the University of Oklahoma (2012-2014). She specializes in the analysis of degraded DNA and proteins, and she led the first metagenomic and metaproteomic investigations of dental calculus. Her research focuses on the study of ancient biomolecules to better understand past human diet, health, and the evolution of the human microbiome, and she has made important contributions to understanding the prehistory of dairying, tracing the evolution of the oral microbiome, and reconstructing the migrations and cultural exchange networks of past human societies.

Kristine Korzow Richter is a Research Associate in the Warinner Lab at Harvard University. She earned her Ph.D. from the Pennsylvania State University in 2014 before completing postdoctoral training in proteomics at the University of York (2015−2018) and the Max Planck Institute for the Science of Human History (2018−2020). She specializes in the preservation, recovery, and identification of ancient proteins for the taxonomic identification of animals. Her research focuses on the study of animal−human interactions including exploitation and management as well as ecosystem reconstruction.

Matthew Collins is a Professor of Biomolecular Archaeology at the Globe Institute of the University of Copenhagen and the MacDonald Chair of Palaeoproteomics at the University of Cambridge. He was awarded a Ph.D. from the University of Glasgow (UCNW) in 1996 and received training in Chemistry and Biochemistry at the University of Leiden (1996−1991) and in microbiology and organic geochemistry at the University of Bristol (1991−1993) before taking a lectureship in Environmental Biogeochemistry at the University of Newcastle. He then moved to the University of York in 2003 to establish BioArCh, a center integrating Biology, Archaeology and Chemistry. In 2016, he was appointed a DNRF Niels Bohr Professor at the University of Copenhagen and the MacDonald Chair of Palaeoproteomics at the University of Cambridge. His research explores the survival of biomolecules, notably proteins in the archaeological and geological record, and the potential to use these to estimate the age and authenticity of samples in an attempt to recover biologically and archaeologically relevant information. A particular interest of his has been the potential of well-dated historical materials, notably parchment and wax, to explore past production processes from farm to factory.

ACKNOWLEDGMENTS

GLOSSARY OF ABBREVIATIONS AND ACRONYMS

AAR = amino acid racemization; the chiral conversion of L-amino acids into D-amino acids

AHSG = α2-HS glycoprotein, also known as fetuin-A; a protein that contributes to biomineralization

AMBN = ameloblastin; a major protein in enamel

AMBP = α-1-Microglobulin/Bikunin Precursor; a protein that has been detected in pottery residues

AMELX = X-chromosome isoform of the amelogenin protein; a major protein in enamel

AMELY = Y-chromosome isoform of the amelogenin protein; a major protein in enamel

AMTN = amelotin; a major protein in enamel

ASPN = asporin; a protein that facilitates tooth attachment to the periodontal ligament

AVD = avidin; a protein present in egg white

b ions = in protein tandem mass spectrometry, b ions are a series of fragment ions that extend from the N-terminus. Low energy collision induced dissociation (CID) typically produces pairs of b ions and y ions by breaking the peptide amide bond.

BGN = biglycan; a protein that facilitates biomineralization

BLASTp = protein–protein basic local alignment search tool; a tool that identifies regions of local similarity between protein sequences; it can be used to infer functional and evolutionary relationships

BLG = β-lactoglobulin; a major protein in the whey fraction of milk

bp = base pairs of double-stranded DNA

BP = Before Present, a standard unit of time used in radiocarbon dating that is calculated as radiocarbon years before 1950. May be uncalibrated (radiocarbon years before present, RCYBP) or calibrated (cal BP). For date estimations obtained without radiometric methods, the units Ka (kiloannuum, thousand years ago) or Ma (megaannuum, million years ago) are recommended.

CERN = Conseil Européen pour la Recherche Nucléaire; the European Organization for Nuclear Research

CID = collision-induced dissociation; a method of precursor ion fragmentation in tandem mass spectrometry that is widely used in proteomics and which primarily produces b and y ions

C-terminus = left-to-right nomenclature of an amino acid chain, referring to the last amino acid in the chain that has a free carboxylic acid group.

C3 = complement component 3; a protein of the innate immune system that plays a key role in the complement system

C18 = octadecyl carbon chain (C18)-bonded silica; used for protein and peptide purification

CBP = corneous β-proteins, formerly known as β-keratins; a group of structural proteins that are the predominant proteins in the hard corneous material of avian and reptilian scales, claws, beaks, and feathers and turtle shells

CD14 = cluster of differentiation 14; a protein of the innate immune system predominantly produced by macrophages that binds bacterial lipopolysaccharide (LPS)

CDS = coding DNA sequence; the portion of a gene that is expressed into protein

CHAD = chondroadherin; a cartilage-associated protein expressed on bone articular surfaces

CHCA = α-cyano-4-hydroxycinnamic acid; a commonly used matrix in MALDI-TOF MS

CLU = clusterin; a widely expressed secretory glycoprotein in mammals, and also in bird eggshell and egg white

COL1 = Type I collagen; the most abundant form of collagen in animals

COL12 = Type XII collagen; a collagen protein that is found in association with type I collagen

cRAP = Common Repository of Adventitious Proteins; a list of common contaminants in mass spectrometry laboratories

CSN1S1 = α S1 casein; a major milk protein

CSN2 = β casein; a major milk protein

CTSG = cathepsin G; a defensive protein produced by cells of the innate immune system, especially neutrophils

D-amino acid = stereoisomeric form of an amino acid in the D-configuration (dextrorotatory, rotates polarized light rightwards). Although present in peptidoglycan and produced by bacteria, D-amino acids contribute minimally to the proteins of most living organisms. Most D-amino acids are believed to form through diagenetic racemization.

Da, kDa = dalton, kilodalton; a dalton is defined as one twelfth the mass of a free neutral atom of ^{12}C at rest

DDA = data-dependent acquisition; a mode of data collection in tandem mass spectrometry in which the most intense precursor ions in a first stage of tandem mass spectrometry are then fragmented and analyzed in a second stage of tandem mass spectrometry

DDJB = DNA Databank of Japan; one of three consortium members in the INSDC

DEFA = defensin α 1, also known as neutrophil defensin 1; a cytotoxic protein produced by cells of the innate immune system, especially neutrophils

DEJ = dentine–enamel junction. The junction between the enamel crown and the underlying tooth dentine; a key site of enamel production during development.

DESI = desorption electrospray ionization

DHB = 2,5-dihydroxybenzoic acid

DIA = data-independent acquisition; a mode of data collection in tandem mass spectrometry in which all precursor ions within a narrow m/z window in a first stage of tandem mass spectrometry are then fragmented in a second stage of mass spectrometry

DISPA = direct infusion shotgun proteome analysis

DNA = deoxyribonucleic acid

DPT = dermatopontin; an extracellular matrix protein that accelerates collagen fibril formation and stabilizes collagen fibrils

EDTA = ethylenediaminetetraacetic acid (EDTA), notable for its 2+ cation chelating ability. Widely used in ancient biomolecular studies to demineralize skeletal remains.

ELISA = enzyme-linked immunosorbent assay; a solid-phase type of immunoassay that can detect protein ligands in solution using antibodies

EMBL-EBI = European Bioinformatics Institute, based in Hinxton, UK; a component of the European Molecular Biology Laboratory, an intragovernmental organization headquartered in Heidelberg, Germany. One of three consortium members in the INSDC and a member of the UniProt Consortium

ENAM = enamelin; a major protein in enamel

Ensembl = a project run by EMBL-EBI that imports primary data from genome and genetic data archive resources and

provides annotation of transcript structures, genomic variants, and regulatory regions

ESI = electrospray ionization; a form of soft ionization used by LC−MS/MS systems

EVA = ethylene-vinyl acetate; an elastomeric polymer used for minimally invasive protein sampling

F2 = coagulation factor II, also known as prothrombin; a protein involved in blood coagulation

F7 = coagulation factor VII; a protein involved in blood coagulation

F9 = coagulation factor IX; a protein involved in blood coagulation

F10 = coagulation factor X; a protein involved in blood coagulation

FASP = filter-aided sample preparation; a method used for protein extraction

FDR = False Discovery Rate; a statistical method for estimating type I errors. FDR-controlling procedures are applied to peptide and protein identifications to minimize spurious results

FT-ICR MS = Fourier-transform ion cyclotron mass spectrometry

GAPB = glyceraldehyde-3-phosphate dehydrogenase; a ubiquitous enzyme involved in glycolysis

GASP = Gel-Aided Sample Preparation; a method used for protein extraction

GenBank = genetic sequence database containing an annotated collection of all publicly available DNA sequences maintained by NCBI

Glu-C = endoproteinase that preferentially cleaves peptide bonds C-terminal to glutamic acid residues; also known as V-8 protease

GLYCAM1 = glycosylation-dependent cell adhesion molecule-1; a mucin-like glycoprotein present in milk

HBA1 = hemoglobin subunit α 1; a major component of hemoglobin in blood

HBBF = fetal hemoglobin subunit beta; a protein differentially expressed in the months before and after birth; relevant for studies animal skins and parchments

HBV = Hepatitis B virus

HOMD = Human Oral Microbiome Database; a curated online database of human oral microbes and associated genomic data and metadata developed and maintained by the Forsyth Institute

HPLC = high performance liquid chromatography; form of chromatographic separation widely used in protein tandem mass spectrometry workflows

HSP90A = inducible cytosolic isoform of heat shock protein 90; a protein differentially expressed in the months before and after birth. Relevant for studies animal skins and parchments.

INSDC = International Nucleotide Sequence Database Collaboration; a global body operated by the EMBL-EBI, NCBI, and DDJB that coordinates the storage and sharing of genetic sequence data, alignments, assemblies, and functional annotations

Ka = kiloannuum; thousand years ago

KAP4-2 = keratin associated protein 4−2; a protein component of wool

kDa = unit of mass corresponding to 1000 Da

α-keratins = Alpha-keratins are a group of structural proteins that are the predominant proteins of vertebrate hair/fur, nails/claws, horns, hooves, quills, and baleen; a minor component of skin

KLK4 = kallikrein related peptidase 4; an enamel protein

KRT75 = Keratin type II cytoskeletal 75; a keratin protein that has been identified within enamel

L-amino acid = stereoisomeric form of an amino acid in the L-configuration (laevorotatory, rotates polarized light leftwards). With few exceptions, proteins within living organisms are made up of L-amino acids.

LC−MS/MS = liquid chromatography tandem mass spectrometry

LEGK = legumin K; an abundant protein in many plant seeds

LESA-MSI = liquid extraction surface analysis mass spectrometry imaging

LPO = lactoperoxidase; a defensive enzyme secreted by mammary and other mucosal glands

LRG1 = leucine rich α-2-glycoprotein 1; a secreted glycoprotein of the innate immune system

LUM = lumican; a protein that supports collagen fibril organization

Lys-C = endoproteinase that cleaves peptide bonds C-terminal to lysine residues

Lys-N = metalloendoprotease that cleaves peptide bonds N-terminal to lysine residues

LYZ = lysozyme; an antimicrobial enzyme that is part of the innate immune system

Ma = megaannuum; million years ago

MALDI-MSI = matrix-assisted laser desorption/ionization mass spectrometry imaging

MALDI-TOF = matrix-assisted laser desorption/ionization time-of-flight mass spectrometry

MGP = matrix gla protein; regulates biomineralization

MMP20 = matrix metalloproteinase-20, also known as enamelysin; an enamel protein

MPO = myeloperoxidase; a prevalent protein in dental calculus that is produced by cells of the innate immune system, especially neutrophils

MRM = multiple reaction monitoring; a method that can be used in the targeted acquisition of tandem mass spectrometry data

MS1 = first mass scan in tandem mass spectrometry

MS2 = second mass scan in tandem mass spectrometry

MSI = mass spectrometry imaging

MUC5B = mucin 5B; a major gel-forming mucin in saliva and other mucus

m/z = mass-to-charge ratio

N-terminus = left-to-right nomenclature of an amino acid chain, referring to the first amino acid in the chain that has a free amine group

NCBI = National Center for Biotechnology Information; a governmental agency based in Bethesda, USA, that develops and coordinates information technology, databases, and software to support research in molecular biology, biochemistry, and genetics. One of three consortium members in the INSDC, it maintains the databases GenBank and RefSeq, among other resources

NCP = noncollagenous protein; refers to the non-collagenous proteins present within predominantly collagenous tissues, such as bone

ODAM = odontogeneic ameloblast-associated protein; a major protein in enamel

OIH = ovoinhibitor; antimicrobial protease inhibitor found in egg white and egg yolk

PCA = principal components analysis

PMF = peptide mass fingerprinting

PIP = prolactin induced protein; a protein involved in immunological function and fluid production

PIR = Protein Information Resource; a UniProt Consortium member

POSTN = periostin; a protein highly expressed in bone periosteum

PRIDE = PRoteomics IDEntifications (PRIDE) database; a major data repository of mass spectrometry-based proteomics data operated by EMBL-EBI

PRM = parallel reaction monitoring; a method that can be used in the targeted acquisition of tandem mass spectrometry data

ProAlanase = endoprotease that preferentially cleaves peptide bonds C-terminal to proline and, to a lesser extent, alanine

PSD = postsource decay

PTM = post-translational modification

PVC = polyvinyl chloride, a synthetic polymer of plastic; used in minimally destructive sampling protocols to obtain protein through the triboelectric effect

R group = a functional group within a molecule that has distinctive chemical properties; the R group of an amino acid determines which amino acid it is

RefSeq = a database of nonredundant annotated sequences representing genomic data, transcripts and proteins maintained by the NCBI

RIA = radioimmunoassay

RNA = ribonucleic acid

RPN2 = ribophorin II; a protein expressed in the rough endoplasmic reticulum; relevant for studies animal skins and parchments

S100A8 = S100 Calcium Binding Protein A8; a protein involved in the regulation of inflammation and immune response

S100A9 = S100 Calcium Binding Protein A9; a protein involved in the regulation of inflammation and immune response

SDS = sodium dodecyl sulfate; a surfactant used during protein extraction

SERPINA1 = Serpin Family A Member 1, also known as α-1-antitrypsin; a serine protease inhibitor

SERPINB14 = ovalbumin; a storage protein that is the most abundant protein in egg white

SERPINF1 = Serpin Family F Member 1; neurotrophic protein, also inhibits angiogenesis

SIB = Swiss Bioinformatics Institute; a member of the UniProt Consortium

SP3 = single-pot, solid-phase-enhanced sample preparation; a method used for protein extraction

SPIN = species by proteome investigation; a DIA workflow for identifying mammalian species using tandem mass spectrometry

SPINK7 = Serine Peptidase Inhibitor Kazal Type 7, also known as ovomucoid; an abundant serine peptidase inhibitor in egg white

SRM = selected reaction monitoring; a synonym for multiple reaction monitoring (MRM)

SwissProt = manually annotated and nonredundant protein sequence database component of the UniProtKB; maintained by the Swiss Institute of Bioinformatics (SIB)

TENP = transiently expressed in neural precursors, also known as BPI fold-containing family B, member 2 and ovoglobulin G2; a major protein in egg white

TF = transferrin; an iron binding transport protein

TfsA = tannerella surface protein A; a major component of the S-layer in Tannerella forsythia, a bacterium associated with dental plaque

TfsB = tannerella surface protein B; a major component of the S-layer in Tannerella forsythia, a bacterium associated with dental plaque

TOF = time-of-flight mass spectrometry

TrEMBL = translated EMBL Nucleotide Sequence Data Library; nonreviewed protein sequences translated from genetic data supplied by EMBL-EBI that have been computationally analyzed and enriched with automatic annotation and classification

UHPLC = ultra high performance liquid chromatography

UMOD = uromodulin; a glycoprotein produced by mammalian kidneys, abundant in urine

UniProtKB = UniProt Knowledgebase; centralized resource for protein metadata, including annotated structural and functional information; maintained by the UniProt consortium, which consists of the EMBL-EBI, SIB, and PIR

VIM = vimentin; a major cytoskeletal protein in mesenchymal cells

VTN = vitronectin; a cell adhesion protein found in serum and tissues

y ions = in protein tandem mass spectrometry, y ions are a series of fragment ions that extend from the C-terminus. Low energy collision induced dissociation (CID) typically produces pairs of b ions and y ions by breaking the peptide amide bond.

ZooMS = Zooarchaeology by Mass Spectrometry; an application of MALDI-TOF collagen PMF for taxonomic identification

REFERENCES

(1) Boyd, W. C.; Boyd, L. G. Blood Grouping Tests on 300 Mummies: With Notes on the Precipitin-Test. *J. Immunol.* **1937**, *32*, 307−319.

(2) Abelson, P. H. Paleobiochemistry: Organic Constituents of Fossils. *Carnegie Institution of Washington, Yearbook* **1954**, *53*, 97−101.

(3) Ostrom, P. H.; Schall, M.; Gandhi, H.; Shen, T.-L.; Hauschka, P. V.; Strahler, J. R.; Gage, D. A. New Strategies for Characterizing Ancient Proteins Using Matrix-Assisted Laser Desorption Ionization Mass Spectrometry. *Geochim. Cosmochim. Acta* **2000**, *64*, 1043−1050.

(4) Hendy, J.; Welker, F.; Demarchi, B.; Speller, C.; Warinner, C.; Collins, M. J. A Guide to Ancient Protein Studies. *Nat. Ecol. Evol.* **2018**, *2*, 791−799.

(5) Hendy, J. Ancient Protein Analysis in Archaeology. *Sci. Adv.* **2021**, *7*, No. eabb9314.

(6) Buckley, M. Paleoproteomics: An Introduction to the Analysis of Ancient Proteins by Soft Ionisation Mass Spectrometry. In *Paleogenomics: Genome-Scale Analysis of Ancient DNA*; Lindqvist, C.; Rajora, O. P., Eds.; Springer International Publishing: Cham, 2019; pp 31−52. DOI: 10.1007/13836_2018_50.

(7) Richter, K. K.; Codlin, M.; Seabrook, M.; Warinner, C. A Primer for ZooMS Applications in Archaeology. *Proc. Natl. Acad. Sci. U.S.A.* **2020**, *119*, No. e2109323119.

(8) Liang, C.; Amelung, W.; Lehmann, J.; Kästner, M. Quantitative Assessment of Microbial Necromass Contribution to Soil Organic Matter. *Glob. Chang. Biol.* **2019**, *25*, 3578−3590.

(9) Demarchi, B.; Hall, S.; Roncal-Herrero, T.; Freeman, C. L.; Woolley, J.; Crisp, M. K.; Wilson, J.; Fotakis, A.; Fischer, R.; Kessler, B. M. Protein Sequences Bound to Mineral Surfaces Persist into Deep Time. *Elife* **2016**, x DOI: 10.7554/eLife.17092.

(10) Rybczynski, N.; Gosse, J. C.; Harington, C. R.; Wogelius, R. A.; Hidy, A. J.; Buckley, M. Mid-Pliocene Warm-Period Deposits in the High Arctic Yield Insight into Camel Evolution. *Nat. Commun.* **2013**, *4*, 1550.

(11) van der Valk, T.; Pečnerová, P.; Díez-Del-Molino, D.; Bergström, A.; Oppenheimer, J.; Hartmann, S.; Xenikoudakis, G.; Thomas, J. A.; Dehasque, M.; Sağlıcan, E.; et al. Million-Year-Old DNA Sheds Light on the Genomic History of Mammoths. *Nature* **2021**, *591*, 265−269.

(12) Orlando, L.; Ginolhac, A.; Zhang, G.; Froese, D.; Albrechtsen, A.; Stiller, M.; Schubert, M.; Cappellini, E.; Petersen, B.; Moltke, I.; et al. Recalibrating Equus Evolution Using the Genome Sequence of an Early Middle Pleistocene Horse. *Nature* **2013**, *499*, 74−78.

(13) Welker, F.; Ramos-Madrigal, J.; Kuhlwilm, M.; Liao, W.; Gutenbrunner, P.; de Manuel, M.; Samodova, D.; Mackie, M.; Allentoft, M. E.; Bacon, A.-M.; et al. Enamel Proteome Shows That Gigantopithecus Was an Early Diverging Pongine. *Nature* **2019**, *576*, 262−265.

(14) 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. Early Pleistocene Enamel Proteome from Dmanisi Resolves Stephanorhinus Phylogeny. *Nature* **2019**, *574*, 103−107.

(15) 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. R. Soc. B: Biol. Sci.* **2017**, *284*, 20170544.

(16) Cleland, T. P.; Schroeter, E. R.; Zamdborg, L.; Zheng, W.; Lee, J. E.; Tran, J. C.; Bern, M.; Duncan, M. B.; Lebleu, V. S.; Ahlf, D. R.; et al. Mass Spectrometry and Antibody-Based Characterization of Blood Vessels from Brachylophosaurus Canadensis. *J. Proteome Res.* **2015**, *14*, 5252−5262.

(17) Schroeter, E. R.; DeHart, C. J.; Cleland, T. P.; Zheng, W.; Thomas, P. M.; Kelleher, N. L.; Bern, M.; Schweitzer, M. H. Expansion for the Brachylophosaurus Canadensis Collagen I Sequence and Additional Evidence of the Preservation of Cretaceous Protein. *J. Proteome Res.* **2017**, *16*, 920−932.

(18) Knoll, A. H. Paleobiological Perspectives on Early Microbial Evolution. *Cold Spring Harb. Perspect. Biol.* **2015**, *7*, a018093.

(19) Baldauf, S. L.; Roger, A. J.; Wenk-Siefert, I.; Doolittle, W. F. A Kingdom-Level Phylogeny of Eukaryotes Based on Combined Protein Data. *Science* **2000**, *290*, 972−977.

(20) 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. Ancient Proteins Resolve the Evolutionary History of Darwin's South American Ungulates. *Nature* **2015**, *522*, 81−84.

(21) Horn, I. R.; Kenens, Y.; Palmblad, N. M.; van der Plas-Duivesteijn, S. J.; Langeveld, B. W.; Meijer, H. J. M.; Dalebout, H.; Marissen, R. J.; Fischer, A.; Vincent Florens, F. B.; et al. Palaeoproteomics of Bird Bones for Taxonomic Classification. *Zool. J. Linn. Soc.* **2019**, *186*, 650−665.

(22) Macek, B.; Forchhammer, K.; Hardouin, J.; Weber-Ban, E.; Grangeasse, C.; Mijakovic, I. Protein Post-Translational Modifications in Bacteria. *Nat. Rev. Microbiol.* **2019**, *17*, 651−664.

(23) Witze, E. S.; Old, W. M.; Resing, K. A.; Ahn, N. G. Mapping Protein Post-Translational Modifications with Mass Spectrometry. *Nat. Methods* **2007**, *4*, 798−806.

(24) Paulus, H. Protein Splicing and Related Forms of Protein Autoprocessing. *Annu. Rev. Biochem.* **2000**, *69*, 447−496.

(25) Vu, L. D.; Gevaert, K.; De Smet, I. Protein Language: Post-Translational Modifications Talking to Each Other. *Trends Plant Sci.* **2018**, *23*, 1068−1080.

(26) Toyama, B. H.; Hetzer, M. W. Protein Homeostasis: Live Long, Won't Prosper. *Nat. Rev. Mol. Cell Biol.* **2013**, *14*, 55−61.

(27) Hochstrasser, M.; Kornitzer, D. Ubiquitin-Dependent Degradation of Transcription Regulators. In *Ubiquitin and the Biology of the Cell*; Peters, J.-M.; Harris, J. R.; Finley, D., Eds.; Springer US: Boston, MA, 1998; pp 279−302. DOI: 10.1007/978-1-4899-1922-9_9.

(28) Fernández-Messina, L.; Reyburn, H. T.; Valés-Gómez, M. A Short Half-Life of ULBP1 at the Cell Surface Due to Internalization and Proteosomal Degradation. *Immunol. Cell Biol.* **2016**, *94*, 479−485.

(29) Helfman, P. M.; Bada, J. L. Aspartic Acid Racemization in Tooth Enamel from Living Humans. *Proc. Natl. Acad. Sci. U. S. A.* **1975**, *72*, 2891−2894.

(30) Stewart, D. N.; Lango, J.; Nambiar, K. P.; Falso, M. J. S.; FitzGerald, P. G.; Rocke, D. M.; Hammock, B. D.; Buchholz, B. A. Carbon Turnover in the Water-Soluble Protein of the Adult Human Lens. *Mol. Vis.* **2013**, *19*, 463−475.

(31) Becker, M. A.; Magoshi, Y.; Sakai, T.; Tuross, N. C. Chemical and Physical Properties of Old Silk Fabrics. *Stud. Conserv.* **1997**, *42*, 27−37.

(32) Good, I. Archaeological Textiles: A Review of Current Research. *Annu. Rev. Anthropol.* **2001**, *30*, 209−226.

(33) Watson, J. D.; Crick, F. H. Molecular Structure of Nucleic Acids; a Structure for Deoxyribose Nucleic Acid. *Nature* **1953**, *171*, 737−738.

(34) Crick, F. H. On Protein Synthesis. *Symp. Soc. Exp. Biol.* **1958**, *12*, 138−163.

(35) Crick, F. Central Dogma of Molecular Biology. *Nature* **1970**, *227*, 561−563.

(36) Boyd, W. C.; Boyd, L. G. An Attempt to Determine the Blood Groups of Mummies. *Proc. Soc. Exp. Biol. Med.* **1934**, *31*, 671−672.

(37) Candela, P. B. Blood-Group Reactions in Ancient Human Skeletons. *Am. J. Phys. Anthropol.* **1936**, *21*, 429−432.

(38) Abelson, P. H. Annual Report of the Director of the Geophysical Laboratory. *Carnegie Institution of Washington Yearbook* **1955**, *54*, 95−152.

(39) Hare, P. E.; Hoering, T. C.; King, K. *Biogeochemistry of Amino Acids*; Wiley, 1980.

(40) Hare, P. E.; Abelson, P. H. Racemization of Amino Acids in Fossil Shells. *Carnegie Institute of Washington Yearbook* **1968**, *66*, 526−528.

(41) Bada, J. L.; Schroeder, R. A. Amino Acid Racemization Reactions and Their Geochemical Implications. *Sci. Nat.* **1975**, *62*, 71−79.

(42) Bada, J. L.; Gillespie, R.; Gowlett, J. A.; Hedges, R. E. Accelerator Mass Spectrometry Radiocarbon Ages of Amino Acid Extracts from Californian Palaeoindian Skeletons. *Nature* **1984**, *312*, 442−444.

(43) Bada, J. L.; Schroeder, R. A.; Carter, G. F. New Evidence for the Antiquity of Man in North America Deduced from Aspartic Acid Racemization. *Science* **1974**, *184*, 791−793.

(44) Wehmiller, J. F. Interlaboratory Comparison of Amino Acid Enantiomeric Ratios in Fossil Pleistocene Mollusks. *Quat. Res.* **1984**, *22*, 109−120.

(45) Wehmiller, J. F.; York, L. L.; Bart, M. L. Amino Acid Racemization Geochronology of Reworked Quaternary Mollusks on U.S. Atlantic Coast Beaches: Implications for Chronostratigraphy, Taphonomy, and Coastal Sediment Transport. *Mar. Geol.* **1995**, *124*, 303−337.

(46) Dickinson, M. R.; Lister, A. M.; Penkman, K. E. H. A New Method for Enamel Amino Acid Racemization Dating: A Closed System Approach. *Quat. Geochronol.* **2019**, *50*, 29−46.

(47) Demarchi, B.; Williams, M. G.; Milner, N.; Russell, N.; Bailey, G.; Penkman, K. Amino Acid Racemization Dating of Marine Shells: A Mound of Possibilities. *Quat. Int.* **2011**, *239*, 114−124.

(48) Kaufman, D. Dating Deep-Lake Sediments by Using Amino Acid Racemization in Fossil Ostracodes. *Geology* **2003**, *31*, 1049−1052.

(49) West, G.; Kaufman, D. S.; Muschitiello, F.; Forwick, M.; Matthiessen, J.; Wollenburg, J.; O'Regan, M. Amino Acid

Racemization in Quaternary Foraminifera from the Yermak Plateau, Arctic Ocean. *Geochronol.* **2019**, *1*, 53–67.

(50) Wyckoff, R. W.; Doberenz, A. R. THE ELECTRON MICROSCOPY OF RANCHO LA BREA BONE. *Proc. Natl. Acad. Sci. U. S. A.* **1965**, *53*, 230–233.

(51) Wyckoff, R. W.; Mccaughey, W. F.; Doberenz, A. R. The Amino Acid Composition of Proteins from Pleistocene Bones. *Biochim. Biophys. Acta* **1964**, *93*, 374–377.

(52) Akiyama, M.; Wyckoff, R. W. The Total Amino Acid Content of Fossil Pecten Shells. *Proc. Natl. Acad. Sci. U. S. A.* **1970**, *67*, 1097–1100.

(53) Miller, M. F., 2nd; Wyckoff, R. W. Proteins in Dinosaur Bones. *Proc. Natl. Acad. Sci. U. S. A.* **1968**, *60*, 176–178.

(54) Doberenz, A. R.; Miller, M. F., 2nd; Wyckoff, R. W. An Analysis of Fossil Enamel Protein. *Calcif. Tissue Res.* **1969**, *3*, 93–95.

(55) de Jong, E. W.; Westbroek, P.; Westbroek, J. W.; Bruning, J. W. Preservation of Antigenic Properties of Macromolecules over 70 Myr. *Nature* **1974**, *252*, 63–64.

(56) Lowenstein, J. M. Immunological Reactions from Fossil Material. *Philos. Trans. R. Soc. London B Biol. Sci.* **1981**, *292*, 143–149.

(57) Lowenstein, J. M.; Sarich, V. M.; Richardson, B. J. Albumin Systematics of the Extinct Mammoth and Tasmanian Wolf. *Nature* **1981**, *291*, 409–411.

(58) Muyzer, G.; Westbroek, P.; De Vrind, J. P. M.; Tanke, J.; Vrijheid, T.; De Jong, E. W.; Bruning, J. W.; Wehmiller, J. F. Immunology and Organic Geochemistry. *Org. Geochem.* **1984**, *6*, 847–855.

(59) Ulrich, M. M.; Perizonius, W. R.; Spoor, C. F.; Sandberg, P.; Vermeer, C. Extraction of Osteocalcin from Fossil Bones and Teeth. *Biochem. Biophys. Res. Commun.* **1987**, *149*, 712–719.

(60) Smith, P. R.; Wilson, M. T. Detection of Haemoglobin in Human Skeletal Remains by ELISA. *J. Archaeol. Sci.* **1990**, *17*, 255–268.

(61) Collins, M. J.; Muyzer, G.; Westbroek, P.; Curry, G. B.; Sandberg, P. A.; Xu, S. J.; Quinn, R.; Mackinnon, D. Preservation of Fossil Biopolymeric Structures: Conclusive Immunological Evidence. *Geochim. Cosmochim. Acta* **1991**, *55*, 2253–2257.

(62) Brandt, E.; Wiechmann, I.; Grupe, G. How Reliable Are Immunological Tools for the Detection of Ancient Proteins in Fossil Bones? *Int. J. Osteoarchaeol.* **2002**, *12*, 307–316.

(63) Leach, J. D. A Brief Comment on the Immunological Identification of Plant Residues on Prehistoric Stone Tools and Ceramics: Results of a Blind Test. *J. Archaeol. Sci.* **1998**, *25*, 171–175.

(64) Schweitzer, M.; Hill, C. L.; Asara, J. M.; Lane, W. S.; Pincus, S. H. Identification of Immunoreactive Material in Mammoth Fossils. *J. Mol. Evol.* **2002**, *55*, 696–705.

(65) Huq, N. L.; Tseng, A.; Chapman, G. E. Partial Amino Acid Sequence of Osteocalcin from an Extinct Species of Ratite Bird. *Biochem. Int.* **1990**, *21*, 491–496.

(66) Smith, J. B. Peptide Sequencing by Edman Degradation. *e LS* **2001**. .

(67) Timp, W.; Timp, G. Beyond Mass Spectrometry, the Next Step in Proteomics. *Sci. Adv.* **2020**, *6*, No. eaax8978.

(68) Swaminathan, J.; Boulgakov, A. A.; Hernandez, E. T.; Bardo, A. M.; Bachman, J. L.; Marotta, J.; Johnson, A. M.; Anslyn, E. V.; Marcotte, E. M. Highly Parallel Single-Molecule Identification of Proteins in Zeptomole-Scale Mixtures. *Nat. Biotechnol.* **2018**, *36*, 1076–1082.

(69) Ostrom, P. H.; Gandhi, H.; Strahler, J. R.; Walker, A. K.; Andrews, P. C.; Leykam, J.; Stafford, T. W.; Kelly, R. L.; Walker, D. N.; Buckley, M.; et al. Unraveling the Sequence and Structure of the Protein Osteocalcin from a 42ka Fossil Horse. *Geochim. Cosmochim. Acta* **2006**, *70*, 2034–2044.

(70) Nielsen-Marsh, C. M.; Ostrom, P. H.; Gandhi, H.; Shapiro, B.; Cooper, A.; Hauschka, P. V.; Collins, M. J. Sequence Preservation of Osteocalcin Protein and Mitochondrial DNA in Bison Bones Older than 55 Ka. *Geology* **2002**, *30*, 1099–1102.

(71) Wu, C.; Duan, J.; Liu, T.; Smith, R. D.; Qian, W.-J. Contributions of Immunoaffinity Chromatography to Deep Proteome Profiling of Human Biofluids. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **2016**, *1021*, 57–68.

(72) Buckley, M.; Collins, M.; Thomas-Oates, J.; Wilson, J. C. Species Identification by Analysis of Bone Collagen Using Matrix-Assisted Laser Desorption/ionisation Time-of-Flight Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2009**, *23*, 3843–3854.

(73) Buckley, M.; Whitcher Kansa, S.; Howard, S.; Campbell, S.; Thomas-Oates, J.; Collins, M. Distinguishing between Archaeological Sheep and Goat Bones Using a Single Collagen Peptide. *J. Archaeol. Sci.* **2010**, *37*, 13–20.

(74) Nielsen-Marsh, C. M.; Richards, M. P.; Hauschka, P. V.; Thomas-Oates, J. E.; Trinkaus, E.; Pettitt, P. B.; Karavanic, I.; Poinar, H.; Collins, M. J. Osteocalcin Protein Sequences of Neanderthals and Modern Primates. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 4409–4413.

(75) Buckley, M.; Anderung, C.; Penkman, K.; Raney, B. J.; Götherström, A.; Thomas-Oates, J.; Collins, M. J. Comparing the Survival of Osteocalcin and mtDNA in Archaeological Bone from Four European Sites. *J. Archaeol. Sci.* **2008**, *35*, 1756–1764.

(76) Buckley, M.; Collins, M.; Thomas-Oates, J. A Method of Isolating the Collagen (I) Alpha2 Chain Carboxytelopeptide for Species Identification in Bone Fragment. *Anal. Biochem.* **2008**, *374*, 325–334.

(77) Buckley, M. Zooarchaeology by Mass Spectrometry (ZooMS) Collagen Fingerprinting for the Species Identification of Archaeological Bone Fragments. In *Zooarchaeology in Practice: Case Studies in Methodology and Interpretation in Archaeofaunal Analysis*; Giovas, C. M., LeFebvre, M. J., Eds.; Springer International Publishing: Cham, 2018; pp 227–247. .

(78) Collins, M.; Buckley, M.; Grundy, H. H.; Thomas-Oates, J.; Wilson, J.; van Doorn, N. ZooMS: The Collagen Barcode and Fingerprints. *Spectrosc. Eur.* **2010**, *22*, 6–10.

(79) Janzen, A.; Richter, K. K.; Mwebi, O.; Brown, S.; Onduso, V.; Gatwiri, F.; Ndiema, E.; Katongo, M.; Goldstein, S. T.; Douka, K.; et al. Distinguishing African Bovids Using Zooarchaeology by Mass Spectrometry (ZooMS): New Peptide Markers and Insights into Iron Age Economies in Zambia. *PLoS One* **2021**, *16*, No. e0251061.

(80) Buckley, M.; Fraser, S.; Herman, J.; Melton, N. D.; Mulville, J.; Pálsdóttir, A. H. Species Identification of Archaeological Marine Mammals Using Collagen Fingerprinting. *J. Archaeol. Sci.* **2014**, *41*, 631–641.

(81) Harvey, V. L.; Daugnora, L.; Buckley, M. Species Identification of Ancient Lithuanian Fish Remains Using Collagen Fingerprinting. *J. Archaeol. Sci.* **2018**, *98*, 102–111.

(82) Eda, M.; Morimoto, M.; Mizuta, T.; Inoué, T. ZooMS for Birds: Discrimination of Japanese Archaeological Chickens and Indigenous Pheasants Using Collagen Peptide Fingerprinting. *J. Archaeol. Sci. Rep.* **2020**, *34*, 102635.

(83) Harvey, V. L.; LeFebvre, M. J.; deFrance, S. D.; Toftgaard, C.; Drosou, K.; Kitchener, A. C.; Buckley, M. Preserved Collagen Reveals Species Identity in Archaeological Marine Turtle Bones from Caribbean and Florida Sites. *R. Soc. Open Sci.* **2019**, *6*, 191137.

(84) Solazzo, C.; Wadsley, M.; Dyer, J. M.; Clerens, S.; Collins, M. J.; Plowman, J. Characterisation of Novel α-Keratin Peptide Markers for Species Identification in Keratinous Tissues Using Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2013**, *27*, 2685–2698.

(85) Hollemeyer, K.; Altmeyer, W.; Heinzle, E. Identification and Quantification of Feathers, Down, and Hair of Avian and Mammalian Origin Using Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry. *Anal. Chem.* **2002**, *74*, 5960–5968.

(86) Presslee, S.; Wilson, J.; Woolley, J.; Best, J.; Russell, D.; Radini, A.; Fischer, R.; Kessler, B.; Boano, R.; Collins, M.; et al. The Identification of Archaeological Eggshell Using Peptide Markers. *Sci. Technol. Archaeol.* **2017**, *3*, 89–99.

(87) 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.* **2013**, *40*, 1797–1804.

(88) Sakalauskaite, J.; Marin, F.; Pergolizzi, B.; Demarchi, B. Shell Palaeoproteomics: First Application of Peptide Mass Fingerprinting

for the Rapid Identification of Mollusc Shells in Archaeology. *J. Proteomics* **2020**, *227*, 103920.

(89) Tokarski, C.; Martin, E.; Rolando, C.; Cren-Olivé, C. Identification of Proteins in Renaissance Paintings by Proteomics. *Anal. Chem.* **2006**, *78*, 1494–1502.

(90) Zubarev, R. A.; Makarov, A. Orbitrap Mass Spectrometry. *Anal. Chem.* **2013**, *85*, 5288–5296.

(91) Cappellini, E.; Jensen, L. J.; Szklarczyk, D.; Ginolhac, A.; da Fonseca, R. A. R.; Stafford, T. W., Jr.; Holen, S. R.; Collins, M. J.; Orlando, L.; Willerslev, E.; et al. Proteomic Analysis of a Pleistocene Mammoth Femur Reveals More than One Hundred Ancient Bone Proteins. *J. Proteome Res.* **2012**, *11*, 917–926.

(92) Warinner, C.; Rodrigues, J. F. M.; Vyas, R.; Trachsel, C.; Shved, N.; Grossmann, J.; Radini, A.; Hancock, Y.; Tito, R. Y.; Fiddyment, S.; et al. Pathogens and Host Immunity in the Ancient Human Oral Cavity. *Nat. Genet.* **2014**, *46*, 336–344.

(93) Buckley, M. Ancient Collagen Reveals Evolutionary History of the Endemic South American "ungulates. *Proceedings of the Royal Society B: Biological Sciences* **2015**, *282*, 20142671.

(94) 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.* **2017**, *114*, 13649–13654.

(95) Shevchenko, A.; Yang, Y.; Knaust, A.; Thomas, H.; Jiang, H.; Lu, E.; Wang, C.; Shevchenko, A. Proteomics Identifies the Composition and Manufacturing Recipe of the 2500-Year Old Sourdough Bread from Subeixi Cemetery in China. *J. Proteomics* **2013**, DOI: 10.1016/j.jprot.2013.11.016.

(96) Bona, A.; Papai, Z.; Maasz, G.; Toth, G. A.; Jambor, E.; Schmidt, J.; Toth, C.; Farkas, C.; Mark, L. Mass Spectrometric Identification of Ancient Proteins as Potential Molecular Biomarkers for a 2000-Year-Old Osteogenic Sarcoma. *PLoS One* **2014**, *9*, No. e87215.

(97) Lluveras-Tenorio, A.; Vinciguerra, R.; Galano, E.; Blaensdorf, C.; Emmerling, E.; Perla Colombini, M.; Birolo, L.; Bonaduce, I. GC/MS and Proteomics to Unravel the Painting History of the Lost Giant Buddhas of Bāmiyān (Afghanistan). *PLoS One* **2017**, *12*, No. e0172990.

(98) Fremout, W.; Kuckova, S.; Crhova, M.; Sanyova, J.; Saverwyns, S.; Hynek, R.; Kodicek, M.; Vandenabeele, P.; Moens, L. Classification of Protein Binders in Artist's Paints by Matrix-Assisted Laser Desorption/Ionisation Time-of-Flight Mass Spectrometry: An Evaluation of Principal Component Analysis (PCA) and Soft Independent Modelling of Class Analogy (SIMCA). *Rapid Commun. Mass Spectrom.* **2011**, *25*, 1631–1640.

(99) Hendy, J.; Colonese, A. C.; Franz, I.; Fernandes, R.; Fischer, R.; Orton, D.; Lucquin, A.; Spindler, L.; Anvari, J.; Stroud, E.; et al. Ancient Proteins from Ceramic Vessels at Çatalhöyük West Reveal the Hidden Cuisine of Early Farmers. *Nat. Commun.* **2018**, *9*, 4064.

(100) Gorski, J. P. Biomineralization of Bone: A Fresh View of the Roles of Non-Collagenous Proteins. *Front. Biosci.* **2011**, *16*, 2598–2621.

(101) Wilkin, S.; Ventresca Miller, A.; Taylor, W. T. T.; Miller, B. K.; Hagan, R. W.; Bleasdale, M.; Scott, A.; Gankhuyg, S.; Ramsøe, A.; Trachsel, C.; et al. Dairy Pastoralism Sustained Eastern Eurasian Steppe Populations for 5000 Years. *Nat. Ecol. Evol.* **2020**, *4*, 346–355.

(102) 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. Exotic Foods Reveal Contact between South Asia and the Near East during the Second Millennium BCE. *Proc. Natl. Acad. Sci. U. S. A.* **2021**, *118*, No. e2014956117.

(103) Drooker, P. B. Introduction. In *Perishable material culture in the Northeast*; Drooker, P. B., Ed.; New York State Museum: Albany, NY, 2004; pp 1–18.

(104) Solazzo, C.; Rogers, P. W.; Weber, L.; Beaubien, H. F.; Wilson, J.; Collins, M. Species Identification by Peptide Mass Fingerprinting (PMF) in Fibre Products Preserved by Association with Copper-Alloy Artefacts. *J. Archaeol. Sci.* **2014**, *49*, 524–535.

(105) Buckley, M.; Melton, N. D.; Montgomery, J. Proteomics Analysis of Ancient Food Vessel Stitching Reveals > 4000-Year-Old Milk Protein. *Rapid Commun. Mass Spectrom.* **2013**, *27*, 531–538.

(106) Brandt, L. Ø.; Schmidt, A. L.; Mannering, U.; Sarret, M.; Kelstrup, C. D.; Olsen, J. V.; Cappellini, E. Species Identification of Archaeological Skin Objects from Danish Bogs: Comparison between Mass Spectrometry-Based Peptide Sequencing and Microscopy-Based Methods. *PLoS One* **2014**, *9*, No. e106875.

(107) Yang, Y.; Shevchenko, A.; Knaust, A.; Abuduresule, I.; Li, W.; Hu, X.; Wang, C.; Shevchenko, A. Proteomics Evidence for Kefir Dairy in Early Bronze Age China. *J. Archaeol. Sci.* **2014**, *45*, 178–186.

(108) Xie, M.; Shevchenko, A.; Wang, B.; Shevchenko, A.; Wang, C.; Yang, Y. Identification of a Dairy Product in the Grass Woven Basket from Gumugou Cemetery (3800 BP, Northwestern China). *Quat. Int.* **2016**, *426*, 158–165.

(109) 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.* **2008**, *22*, 2751–2767.

(110) Barker, A.; Dombrosky, J.; Venables, B.; Wolverton, S. Taphonomy and Negative Results: An Integrated Approach to Ceramic-Bound Protein Residue Analysis. *J. Archaeol. Sci.* **2018**, *94*, 32–43.

(111) Eisele, J. A.; Fowler, D. D.; Haynes, G.; Lewis, R. A. Survival and Detection of Blood Residues on Stone Tools. *Antiquity* **1995**, *69*, 36–46.

(112) Leach, J. D.; Mauldin, R. P. Additional Comments on Blood Residue Analysis in Archaeology. *Antiquity* **1995**, *69*, 1020–1022.

(113) Cattaneo, C.; Gelsthorpe, K.; Phillips, P.; Sokol, R. J. Blood Residues on Stone Tools: Indoor and Outdoor Experiments. *World Archaeol.* **1993**, *25*, 29–43.

(114) Covington, A. D. Modern Tanning Chemistry. *Chem. Soc. Rev.* **1997**, *26*, 111–126.

(115) Yu, T.-Y.; Morton, J. D.; Clerens, S.; Dyer, J. M. Cooking-Induced Protein Modifications in Meat. *Compr. Rev. Food Sci.* **2017**, *16*, 141–159.

(116) Gil-Bona, A.; Bidlack, F. B. Tooth Enamel and Its Dynamic Protein Matrix. *Int. J. Mol. Sci.* **2020**, *21*, 4458.

(117) Hedges, J. I. Global Giogeochemical Cycles: Progress and Problems. *Mar. Chem.* **1992**, *39*, 67–93.

(118) Benbow, M. E.; Barton, P. S.; Ulyshen, M. D.; Beasley, J. C.; DeVault, T. L.; Strickland, M. S.; Tomberlin, J. K.; Jordan, H. R.; Pechal, J. L. Necrobiome Framework for Bridging Decomposition Ecology of Autotrophically and Heterotrophically Derived Organic Matter. *Ecol. Monogr.* **2019**, *89*, No. e01331.

(119) Solazzo, C.; Dyer, J. M.; Clerens, S.; Plowman, J.; Peacock, E. E.; Collins, M. J. Proteomic Evaluation of the Biodegradation of Wool Fabrics in Experimental Burials. *Int. Biodeterior. Biodegradation* **2013**, *80*, 48–59.

(120) Saitta, E. T.; Rogers, C.; Brooker, R. A.; Abbott, G. D.; Kumar, S.; O'Reilly, S. S.; Donohoe, P.; Dutta, S.; Summons, R. E.; Vinther, J. Low Fossilization Potential of Keratin Protein Revealed by Experimental Taphonomy. *Palaeontology* **2017**, *60*, 547–556.

(121) Wadsworth, C.; Buckley, M. Proteome Degradation in Fossils: Investigating the Longevity of Protein Survival in Ancient Bone. *Rapid Commun. Mass Spectrom.* **2014**, *28*, 605–615.

(122) Brownlow, S.; Morais Cabral, J. H.; Cooper, R.; Flower, D. R.; Yewdall, S. J.; Polikarpov, I.; North, A. C. T.; Sawyer, L. Bovine β-Lactoglobulin at 1.8 Å Resolution — Still an Enigmatic Lipocalin. *Structure* **1997**, *5*, 481–495.

(123) Pekar, J.; Ret, D.; Untersmayr, E. Stability of Allergens. *Mol. Immunol.* **2018**, *100*, 14–20.

(124) Procopio, N.; Williams, A.; Chamberlain, A. T.; Buckley, M. Forensic Proteomics for the Evaluation of the Post-Mortem Decay in Bones. *J. Proteomics* **2018**, *177*, 21–30.

(125) Kendall, C.; Eriksen, A. M. H.; Kontopoulos, I.; Collins, M. J.; Turner-Walker, G. Diagenesis of Archaeological Bone and Tooth. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **2018**, *491*, 21–37.

(126) Demarchi, B. *Amino Acids and Proteins in Fossil Biominerals: An Introduction for Archaeologists and Palaeontologists*; John Wiley & Sons, 2020. DOI: 10.1002/9781119089537.

(127) Cleland, T. P.; Schroeter, E. R.; Colleary, C. Diagenetiforms: A New Term to Explain Protein Changes as a Result of Diagenesis in Paleoproteomics. *J. Proteomics* **2021**, *230*, 103992.

(128) Cleland, T. P.; Schroeter, E. R.; Schweitzer, M. H. Biologically and Diagenetically Derived Peptide Modifications in Moa Collagens. *Proc. Biol. Sci.* **2015**, *282*, 20150015.

(129) Oldenburg, T.; Brown, M.; Inwood, J.; Radović, J.; Snowdon, R.; Larter, S.; Mercader, J. A Novel Route for Identifying Starch Diagenetic Products in the Archaeological Record. *PLoS One* **2021**, *16*, No. e0258779.

(130) Collins, M. J.; Bishop, A. N.; Farrimond, P. Sorption by Mineral Surfaces: Rebirth of the Classical Condensation Pathway for Kerogen Formation? *Geochim. Cosmochim. Acta* **1995**, *59*, 2387−2391.

(131) Walton, D. Degradation of Intracrystalline Proteins and Amino Acids in Fossil Brachiopods. *Org. Geochem.* **1998**, *28*, 389−410.

(132) Buckley, M.; Wadsworth, C. Proteome Degradation in Ancient Bone: Diagenesis and Phylogenetic Potential. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **2014**, *416*, 69−79.

(133) Craig, O. E.; Collins, M. J. The Removal of Protein from Mineral Surfaces: Implications for Residue Analysis of Archaeological Materials. *J. Archaeol. Sci.* **2002**, *29*, 1077−1082.

(134) Carrera, M. Proteomics and Food Analysis: Principles, Techniques, and Applications. *Foods* **2021**, *10*, 2538.

(135) Keller, B. O.; Sui, J.; Young, A. B.; Whittal, R. M. Interferences and Contaminants Encountered in Modern Mass Spectrometry. *Anal. Chim. Acta* **2008**, *627*, 71−81.

(136) Schroeter, E. R.; DeHart, C. J.; Schweitzer, M. H.; Thomas, P. M.; Kelleher, N. L. Bone Protein "Extractomics": Comparing the Efficiency of Bone Protein Extractions of *Gallus Gallus* in Tandem Mass Spectrometry, with an Eye Towards Paleoproteomics. *PeerJ* **2016**, *4*, No. e2603.

(137) 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. Palaeoproteomic Profiling of Conservation Layers on a 14th Century Italian Wall Painting. *Angew. Chem., Int. Ed. Engl.* **2018**, *57*, 7369−7374.

(138) Cleland, T. P.; Voegele, K.; Schweitzer, M. H. Empirical Evaluation of Bone Extraction Protocols. *PLoS One* **2012**, *7*, No. e31443.

(139) Wang, N.; Brown, S.; Ditchfield, P.; Hebestreit, S.; Kozilikin, M.; Luu, S.; Wedage, O.; Grimaldi, S.; Chazan, M.; Kolska, L. H.; et al. Testing the Efficacy and Comparability of ZooMS Protocols on Archaeological Bone. *J. Proteomics* **2021**, *233*, 104078.

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

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

(142) Procopio, N.; Hopkins, R. J. A.; Harvey, V. L.; Buckley, M. Proteome Variation with Collagen Yield in Ancient Bone. *J. Proteome Res.* **2021**, *20*, 1754−1769.

(143) Cleland, T. P.; Sarancha, J. J.; France, C. A. M. Proteomic Profile of Bone "Collagen" Extracted for Stable Isotopes: Implications for Bulk and Single Amino Acid Analyses. *Rapid Commun. Mass Spectrom.* **2021**, *35*, No. e9025.

(144) Charlton, S.; Alexander, M.; Collins, M.; Milner, N.; Mellars, P.; O'Connell, T. C.; Stevens, R. E.; Craig, O. E. Finding Britain's Last Hunter-Gatherers: A New Biomolecular Approach to "Unidentifiable" Bone Fragments Utilising Bone Collagen. *J. Archaeol. Sci.* **2016**, *73*, 55−61.

(145) Fagernäs, Z.; García-Collado, M. I.; Hendy, J.; Hofman, C. A.; Speller, C.; Velsko, I.; Warinner, C. A Unified Protocol for Simultaneous Extraction of DNA and Proteins from Archaeological Dental Calculus. *J. Archaeol. Sci.* **2020**, *118*, 105135.

(146) Du, J.; Zhu, Z.; Yang, J.; Wang, J.; Jiang, X. A Comparative Study on the Extraction Effects of Common Agents on Collagen-Based Binders in Mural Paintings. *Herit. Sci.* **2021**, *9*, 45.

(147) Wiśniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M. Universal Sample Preparation Method for Proteome Analysis. *Nat. Methods* **2009**, *6*, 359−362.

(148) 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* **2015**, *15*, 1224−1229.

(149) Hendy, J.; Warinner, C.; Bouwman, A.; Collins, M. J.; Fiddyment, S.; Fischer, R.; Hagan, R.; Hofman, C. A.; Holst, M.; Chaves, E. Proteomic Evidence of Dietary Sources in Ancient Dental Calculus. *Proc. Biol. Sci.* **2018**, DOI: 10.1098/rspb.2018.0977.

(150) Hughes, C. S.; Foehr, S.; Garfield, D. A.; Furlong, E. E.; Steinmetz, L. M.; Krijgsveld, J. Ultrasensitive Proteome Analysis Using Paramagnetic Bead Technology. *Mol. Syst. Biol.* **2014**, *10*, 757.

(151) Cleland, T. P. Human Bone Paleoproteomics Utilizing the Single-Pot, Solid-Phase-Enhanced Sample Preparation Method to Maximize Detected Proteins and Reduce Humics. *J. Proteome Res.* **2018**, *17*, 3976−3983.

(152) Cleland, T. P. Solid Digestion of Demineralized Bone as a Method To Access Potentially Insoluble Proteins and Post-Translational Modifications. *J. Proteome Res.* **2018**, *17*, 536−542.

(153) Jersie-Christensen, R. R.; Sultan, A.; Olsen, J. V. Simple and Reproducible Sample Preparation for Single-Shot Phosphoproteomics with High Sensitivity. *Methods Mol. Biol.* **2016**, *1355*, 251−260.

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

(155) 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.* **2011**, *3*, 281−289.

(156) Fiddyment, S.; Holsinger, B.; Ruzzier, C.; Devine, A.; Binois, A.; Albarella, U.; Fischer, R.; Nichols, E.; Curtis, A.; Cheese, E.; et al. Animal Origin of 13th-Century Uterine Vellum Revealed Using Noninvasive Peptide Fingerprinting. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 15066−15071.

(157) Manfredi, M.; Barberis, E.; Gosetti, F.; Conte, E.; Gatti, G.; Mattu, C.; Robotti, E.; Zilberstein, G.; Koman, I.; Zilberstein, S.; Marengo, E.; Righetti, P. G. Method for Noninvasive Analysis of Proteins and Small Molecules from Ancient Objects. *Anal. Chem.* **2017**, *89*, 3310−3317.

(158) Cleland, T. P.; Vashishth, D. Bone Protein Extraction without Demineralization Using Principles from Hydroxyapatite Chromatography. *Anal. Biochem.* **2015**, *472*, 62−66.

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

(160) Ebsen, J. A.; Haase, K.; Larsen, R.; Sommer, D. V. P.; Brandt, L. Ø. Identifying Archaeological Leather - Discussing the Potential of Grain Pattern Analysis and Zooarchaeology by Mass Spectrometry (ZooMS) through a Case Study Involving Medieval Shoe Parts from Denmark. *J. Cult. Herit.* **2019**, *39*, 21−31.

(161) Szpak, P.; Krippner, K.; Richards, M. P. Effects of Sodium Hydroxide Treatment and Ultrafiltration on the Removal of Humic Contaminants from Archaeological Bone. *Int. J. Osteoarchaeol.* **2017**, *27*, 1070−1077.

(162) Oonk, S.; Cappellini, E.; Collins, M. J. Soil Proteomics: An Assessment of Its Potential for Archaeological Site Interpretation. *Org. Geochem.* **2012**, *50*, 57−67.

(163) Salamon, M.; Tuross, N.; Arensburg, B.; Weiner, S. Relatively Well Preserved DNA Is Present in the Crystal Aggregates of Fossil Bones. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 13783−13788.

(164) Zhang, Y.; Fonslow, B. R.; Shan, B.; Baek, M.-C.; Yates, J. R., 3rd. Protein Analysis by Shotgun/Bottom-up Proteomics. *Chem. Rev.* **2013**, *113*, 2343−2394.

(165) Laskay, Ü. A.; Lobas, A. A.; Srzentić, K.; Gorshkov, M. V.; Tsybin, Y. O. Proteome Digestion Specificity Analysis for Rational Design of Extended Bottom-up and Middle-down Proteomics Experiments. *J. Proteome Res.* **2013**, *12*, 5558−5569.

(166) Swaney, D. L.; Wenger, C. D.; Coon, J. J. Value of Using Multiple Proteases for Large-Scale Mass Spectrometry-Based Proteomics. *J. Proteome Res.* **2010**, *9*, 1323−1329.

(167) Samodova, D.; Hosfield, C. M.; Cramer, C. N.; Giuli, M. V.; Cappellini, E.; Franciosa, G.; Rosenblatt, M. M.; Kelstrup, C. D.; Olsen, J. V. ProAlanase Is an Effective Alternative to Trypsin for Proteomics Applications and Disulfide Bond Mapping. *Mol. Cell. Proteomics* **2020**, *19*, 2139−2157.

(168) Lanigan, L. T.; Mackie, M.; Feine, S.; Hublin, J.-J.; Schmitz, R. W.; Wilcke, A.; Collins, M. J.; Cappellini, E.; Olsen, J. V.; Taurozzi, A. J.; et al. Multi-Protease Analysis of Pleistocene Bone Proteomes. *J. Proteomics* **2020**, *228*, 103889.

(169) Calvano, C. D.; Rigante, E. C. L.; Cataldi, T. R. I.; Sabbatini, L. In Situ Hydrogel Extraction with Dual-Enzyme Digestion of Proteinaceous Binders: The Key for Reliable Mass Spectrometry Investigations of Artworks. *Anal. Chem.* **2020**, *92*, 10257−10261.

(170) Bédouet, L.; Marie, A.; Berland, S.; Marie, B.; Auzoux-Bordenave, S.; Marin, F.; Milet, C. Proteomic Strategy for Identifying Mollusc Shell Proteins Using Mild Chemical Degradation and Trypsin Digestion of Insoluble Organic Shell Matrix: A Pilot Study on *Haliotis Tuberculata. Mar. Biotechnol.* **2012**, *14*, 446−458.

(171) Stewart, N. A.; Molina, G. F.; Issa, J. P. M.; Yates, N. A.; Sosovicka, M.; Vieira, A. R.; Line, S. R. P.; Montgomery, J.; Gerlach, R. F. The Identification of Peptides by nanoLC−MS/MS from Human Surface Tooth Enamel Following a Simple Acid Etch Extraction. *RSC Adv.* **2016**, *6*, 61673−61679.

(172) 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. The Dental Proteome of Homo Antecessor. *Nature* **2020**, *580*, 235−238.

(173) Toby, T. K.; Fornelli, L.; Kelleher, N. L. Progress in Top-Down Proteomics and the Analysis of Proteoforms. *Annu. Rev. Anal. Chem.* **2016**, *9*, 499−519.

(174) Melby, J. A.; Roberts, D. S.; Larson, E. J.; Brown, K. A.; Bayne, E. F.; Jin, S.; Ge, Y. Novel Strategies to Address the Challenges in Top-down Proteomics. *J. Am. Soc. Mass Spectrom.* **2021**, *32*, 1278−1294.

(175) Aebersold, R.; Goodlett, D. R. Mass Spectrometry in Proteomics. *Chem. Rev.* **2001**, *101*, 269−295.

(176) Karas, M.; Bachmann, D.; Bahr, U.; Hillenkamp, F. Matrix-Assisted Ultraviolet Laser Desorption of Non-Volatile Compounds. *Int. J. Mass Spectrom. Ion Process.* **1987**, *78*, 53−68.

(177) Karas, M.; Hillenkamp, F. Laser Desorption Ionization of Proteins with Molecular Masses Exceeding 10,000 Da. *Anal. Chem.* **1988**, *60*, 2299−2301.

(178) Brown, S.; Douka, K.; Collins, M. J.; Richter, K. K. On the Standardization of ZooMS Nomenclature. *J. Proteomics* **2021**, *235*, 104041.

(179) Stover, D. A.; Verrelli, B. C. Comparative Vertebrate Evolutionary Analyses of Type I Collagen: Potential of COL1a1 Gene Structure and Intron Variation for Common Bone-Related Diseases. *Mol. Biol. Evol.* **2011**, *28*, 533−542.

(180) Boot-Handford, R. P.; Tuckwell, D. S. Fibrillar Collagen: The Key to Vertebrate Evolution? A Tale of Molecular Incest. *Bioessays* **2003**, *25*, 142−151.

(181) Wada, H.; Okuyama, M.; Satoh, N.; Zhang, S. Molecular Evolution of Fibrillar Collagen in Chordates, with Implications for the Evolution of Vertebrate Skeletons and Chordate Phylogeny. *Evol. Dev.* **2006**, *8*, 370−377.

(182) Korzow Richter, K.; McGrath, K.; Masson-MacLean, E.; Hickinbotham, S.; Tedder, A.; Britton, K.; Bottomley, Z.; Dobney, K.; Hulme-Beaman, A.; Zona, M.; Fischer, R.; Collins, M. J.; Speller, C. F.; et al. What's the Catch? Archaeological Application of Rapid Collagen-Based Species Identification for Pacific Salmon. *J. Archaeol. Sci.* **2020**, *116*, 105116.

(183) Wagner, A.; Richter, K. K.; Ludes, E.; Arbogast, R.-M.; Carita, D.; Guidez, A.; Soussoko, S.; Boivin, N. L.; Marche, J.-C.; Wandhammer, M.-D.; et al. Whale Bone Puzzles: Reconstructing and Identifying Historical Whale Skeletons Using Archive Records, Osteology, and Zooarchaeology by Mass Spectrometry (ZooMS). *J. Conserv. Mus. Stud.* **2020**, *18*, 1−12.

(184) Douka, K.; Brown, S.; Higham, T.; Pääbo, S.; Derevianko, A.; Shunkov, M. FINDER Project: Collagen Fingerprinting (ZooMS) for the Identification of New Human Fossils. *Antiquity* **2019**, DOI: 10.15184/aqy.2019.3.

(185) Brown, S.; Wang, N.; Oertle, A.; Kozlikin, M. B.; Shunkov, M. V.; Derevianko, A. P.; Comeskey, D.; Jope-Street, B.; Harvey, V. L.; Chowdhury, M. P.; et al. Zooarchaeology through the Lens of Collagen Fingerprinting at Denisova Cave. *Sci. Rep.* **2021**, *11*, 1−10.

(186) Tran, N. H.; Qiao, R.; Xin, L.; Chen, X.; Liu, C.; Zhang, X.; Shan, B.; Ghodsi, A.; Li, M. Deep Learning Enables De Novo Peptide Sequencing from Data-Independent-Acquisition Mass Spectrometry. *Nat. Methods* **2019**, *16*, 63−66.

(187) Yates, J. R., III. A Century of Mass Spectrometry: From Atoms to Proteomes. *Nat. Methods* **2011**, *8*, 633−637.

(188) Eliuk, S.; Makarov, A. Evolution of Orbitrap Mass Spectrometry Instrumentation. *Annu. Rev. Anal. Chem.* **2015**, *8*, 61−80.

(189) Asara, J. M.; Schweitzer, M. H.; Freimark, L. M.; Phillips, M.; Cantley, L. C. Protein Sequences from Mastodon and *Tyrannosaurus Rex* Revealed by Mass Spectrometry. *Science* **2007**, *316*, 280−285.

(190) Pfammatter, S.; Bonneil, E.; McManus, F. P.; Prasad, S.; Bailey, D. J.; Belford, M.; Dunyach, J.-J.; Thibault, P. A Novel Differential Ion Mobility Device Expands the Depth of Proteome Coverage and the Sensitivity of Multiplex Proteomic Measurements. *Mol. Cell. Proteomics* **2018**, *17*, 2051−2067.

(191) Shishkova, E.; Hebert, A. S.; Coon, J. J. Now, More Than Ever, Proteomics Needs Better Chromatography. *Cell Syst.* **2016**, *3*, 321−324.

(192) Mikhailov, V. A.; Griffiths, R. L.; Cooper, H. J. Liquid Extraction Surface Analysis for Native Mass Spectrometry: Protein Complexes and Ligand Binding. *Int. J. Mass Spectrom.* **2017**, *420*, 43−50.

(193) Takáts, Z.; Wiseman, J. M.; Cooks, R. G. Ambient Mass Spectrometry Using Desorption Electrospray Ionization (DESI): Instrumentation, Mechanisms and Applications in Forensics, Chemistry, and Biology. *J. Mass Spectrom.* **2005**, *40*, 1261−1275.

(194) Hale, O. J.; Cooper, H. J. In Situ Mass Spectrometry Analysis of Intact Proteins and Protein Complexes from Biological Substrates. *Biochem. Soc. Trans.* **2020**, *48*, 317−326.

(195) Kocurek, K. I.; Griffiths, R. L.; Cooper, H. J. Ambient Ionisation Mass Spectrometry for in Situ Analysis of Intact Proteins. *J. Mass Spectrom.* **2018**, *53*, 565−578.

(196) Welker, F.; Smith, G. M.; Hutson, J. M.; Kindler, L.; Garcia-Moreno, A.; Villaluenga, A.; Turner, E.; Gaudzinski-Windheuser, S. Middle Pleistocene Protein Sequences from the Rhinoceros Genus *Stephanorhinus* and the Phylogeny of Extant and Extinct Middle/Late Pleistocene Rhinocerotidae. *PeerJ* **2017**, *5*, No. e3033.

(197) Sakalauskaite, J.; Andersen, S. H.; Biagi, P.; Borrello, M. A.; Cocquerez, T.; Colonese, A. C.; Dal Bello, F.; Girod, A.; Heumüller, M.; Koon, H. "Palaeoshellomics" Reveals the Use of Freshwater Mother-of-Pearl in Prehistory. *Elife* **2019**, DOI: 10.7554/eLife.45644.

(198) Wasinger, V. C.; Curnoe, D.; Bustamante, S.; Mendoza, R.; Shoocongdej, R.; Adler, L.; Baker, A.; Chintakanon, K.; Boel, C.; Tacon, P. S. C. Analysis of the Preserved Amino Acid Bias in Peptide Profiles of Iron Age Teeth from a Tropical Environment Enable Sexing of Individuals Using Amelogenin MRM. *Proteomics* **2019**, *19*, No. e1800341.

(199) Vinciguerra, R.; Illiano, A.; De Chiaro, A.; Carpentieri, A.; Lluveras-Tenorio, A.; Bonaduce, I.; Marino, G.; Pucci, P.; Amoresano, A.; Birolo, L. Identification of Proteinaceous Binders in Paintings: A Targeted Proteomic Approach for Cultural Heritage. *Microchem. J.* **2019**, *144*, 319−328.

(200) Degano, I.; Biesaga, M.; Colombini, M. P.; Trojanowicz, M. Historical and Archaeological Textiles: An Insight on Degradation Products of Wool and Silk Yarns. *J. Chromatogr. A* **2011**, *1218*, 5837−5847.

(201) Kumazawa, Y.; Taga, Y.; Takashima, M.; Hattori, S. A Novel LC−MS Method Using Collagen Marker Peptides for Species Identification of Glue Applicable to Samples with Multiple Animal Origins. *Herit. Sci.* **2018**, *6*, 43.

(202) Froment, C.; Hourset, M.; Sáenz-Oyhéréguy, N.; Mouton-Barbosa, E.; Willmann, C.; Zanolli, C.; Esclassan, R.; Donat, R.; Thèves, C.; Burlet-Schiltz, O.; et al. Analysis of 5000 Year-Old Human Teeth Using Optimized Large-Scale and Targeted Proteomics Approaches for Detection of Sex-Specific Peptides. *J. Proteomics* **2020**, *211*, 103548.

(203) Tiwary, S.; Levy, R.; Gutenbrunner, P.; Salinas Soto, F.; Palaniappan, K. K.; Deming, L.; Berndl, M.; Brant, A.; Cimermancic, P.; Cox, J. High-Quality MS/MS Spectrum Prediction for Data-Dependent and Data-Independent Acquisition Data Analysis. *Nat. Methods* **2019**, *16*, 519−525.

(204) Doerr, A. DIA Mass Spectrometry. *Nat. Methods* **2015**, *12*, 35−35.

(205) 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. SPIN Enables High Throughput Species Identification of Archaeological Bone by Proteomics. *Nat. Commun.* **2022**, *13*, 2458.

(206) Strohalm, M.; Hassman, M.; Kosata, B.; Kodíček, M. mMass Data Miner: An Open Source Alternative for Mass Spectrometric Data Analysis. *Rapid Commun. Mass Spectrom.* **2008**, *22*, 905−908.

(207) Gu, M.; Buckley, M. Semi-Supervised Machine Learning for Automated Species Identification by Collagen Peptide Mass Fingerprinting. *BMC Bioinform.* **2018**, *19*, 1−9.

(208) Hickinbotham, S.; Fiddyment, S.; Stinson, T. L.; Collins, M. J. How to Get Your Goat: Automated Identification of Species from MALDI-ToF Spectra. *Bioinformatics* **2020**, *36*, 3719−3725.

(209) Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S. Probability-Based Protein Identification by Searching Sequence Databases Using Mass Spectrometry Data. *Electrophoresis* **1999**, *20*, 3551−3567.

(210) Cox, J.; Mann, M. MaxQuant Enables High Peptide Identification Rates, Individualized P.p.b.-Range Mass Accuracies and Proteome-Wide Protein Quantification. *Nat. Biotechnol.* **2008**, *26*, 1367−1372.

(211) Eng, J. K.; McCormack, A. L.; Yates, J. R. An Approach to Correlate Tandem Mass Spectral Data of Peptides with Amino Acid Sequences in a Protein Database. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 976−989.

(212) Ma, B.; Zhang, K.; Hendrie, C.; Liang, C.; Li, M.; Doherty-Kirby, A.; Lajoie, G. PEAKS: Powerful Software for Peptide De Novo Sequencing by Tandem Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2003**, *17*, 2337−2342.

(213) Bern, M.; Kil, Y. J.; Becker, C. Byonic: Advanced Peptide and Protein Identification Software. *Curr. Protoc. Bioinform.* **2012**, *40*, 13−20.

(214) Searle, B. C. Scaffold: A Bioinformatic Tool for Validating MS/MS-Based Proteomic Studies. *Proteomics* **2010**, *10*, 1265−1269.

(215) Pruitt, K. D.; Tatusova, T.; Maglott, D. R. NCBI Reference Sequence (RefSeq): A Curated Non-Redundant Sequence Database of Genomes, Transcripts and Proteins. *Nucleic Acids Res.* **2004**, *33*, D501−D504.

(216) O'Leary, N. A.; Wright, M. W.; Brister, J. R.; Ciufo, S.; Haddad, D.; McVeigh, R.; Rajput, B.; Robbertse, B.; Smith-White, B.; Ako-Adjei, D.; et al. Reference Sequence (RefSeq) Database at NCBI:

Current Status, Taxonomic Expansion, and Functional Annotation. *Nucleic Acids Res.* **2016**, *44*, D733−D745.

(217) Howe, K. L.; Achuthan, P.; Allen, J.; Allen, J.; Alvarez-Jarreta, J.; Amode, M. R.; Armean, I. M.; Azov, A. G.; Bennett, R.; Bhai, J.; et al. Ensembl 2021. *Nucleic Acids Res.* **2021**, *49*, D884−D891.

(218) Hubbard, T.; Barker, D.; Birney, E.; Cameron, G.; Chen, Y.; Clark, L.; Cox, T.; Cuff, J.; Curwen, V.; Down, T.; et al. The Ensembl Genome Database Project. *Nucleic Acids Res.* **2002**, *30*, 38−41.

(219) Harris, T. W.; Antoshechkin, I.; Bieri, T.; Blasiar, D.; Chan, J.; Chen, W. J.; De La Cruz, N.; Davis, P.; Duesbury, M.; Fang, R.; et al. WormBase: A Comprehensive Resource for Nematode Research. *Nucleic Acids Res.* **2010**, *38*, D463−D467.

(220) Howe, K. L.; Bolt, B. J.; Shafie, M.; Kersey, P.; Berriman, M. WormBase ParaSite - A Comprehensive Resource for Helminth Genomics. *Mol. Biochem. Parasitol.* **2017**, *215*, 2−10.

(221) Lawson, D.; Arensburger, P.; Atkinson, P.; Besansky, N. J.; Bruggner, R. V.; Butler, R.; Campbell, K. S.; Christophides, G. K.; Christley, S.; Dialynas, E.; et al. VectorBase: A Data Resource for Invertebrate Vector Genomics. *Nucleic Acids Res.* **2009**, *37*, D583−D587.

(222) Chen, T.; Yu, W. H.; Izard, J.; Baranova, O. V.; Lakshmanan, A.; Dewhirst, F. E. The Human Oral Microbiome Database: A Web Accessible Resource for Investigating Oral Microbe Taxonomic and Genomic Information. *Database* **2010**, *2010*, baq013.

(223) Bleasdale, M.; Richter, K. K.; Janzen, A.; Brown, S.; Scott, A.; Zech, J.; Wilkin, S.; Wang, K.; Schiffels, S.; Desideri, J.; et al. Ancient Proteins Provide Evidence of Dairy Consumption in Eastern Africa. *Nat. Commun.* **2021**, *12*, 632.

(224) Kumar, D.; Yadav, A. K.; Dash, D. Choosing an Optimal Database for Protein Identification from Tandem Mass Spectrometry Data. *Methods Mol. Biol.* **2017**, *1549*, 17−29.

(225) Levitsky, L. I.; Ivanov, M. V.; Lobas, A. A.; Gorshkov, M. V. Unbiased False Discovery Rate Estimation for Shotgun Proteomics Based on the Target-Decoy Approach. *J. Proteome Res.* **2017**, *16*, 393−397.

(226) Buckley, M.; Lawless, C.; Rybczynski, N. Collagen Sequence Analysis of Fossil Camels, *Camelops* and C.f. *Paracamelus*, from the Arctic and Sub-Arctic of Plio-Pleistocene North America. *J. Proteomics* **2019**, *194*, 218−225.

(227) Muth, T.; Hartkopf, F.; Vaudel, M.; Renard, B. Y. A Potential Golden Age to Come - Current Tools, Recent Use Cases, and Future Avenues for De Novo Sequencing in Proteomics. *Proteomics* **2018**, *18*, 1700150.

(228) Gutenbrunner, P.; Kyriakidou, P.; Welker, F.; Cox, J. Spectrum Graph-Based De Novo Sequencing Algorithm MaxNovo Achieves High Peptide Identification Rates in Collisional Dissociation MS/MS Spectra. *bioRxiv* **2021**, 2021.09.04.458985.

(229) Buckley, M.; Larkin, N.; Collins, M. Mammoth and Mastodon Collagen Sequences; Survival and Utility. *Geochim. Cosmochim. Acta* **2011**, *75*, 2007−2016.

(230) Gurdeep Singh, R.; Tanca, A.; Palomba, A.; Van der Jeugt, F.; Verschaffelt, P.; Uzzau, S.; Martens, L.; Dawyndt, P.; Mesuere, B. Unipept 4.0: Functional Analysis of Metaproteome Data. *J. Proteome Res.* **2019**, *18*, 606−615.

(231) Knudsen, G. M.; Chalkley, R. J. The Effect of Using an Inappropriate Protein Database for Proteomic Data Analysis. *PLoS One* **2011**, *6*, No. e20873.

(232) Verheggen, K.; Raeder, H.; Berven, F. S.; Martens, L.; Barsnes, H.; Vaudel, M. Anatomy and Evolution of Database Search Engines - A Central Component of Mass Spectrometry Based Proteomic Workflows. *Mass Spectrom. Rev.* **2020**, *39*, 292−306.

(233) 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* **2017**, *152*, 102−108.

(234) 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**. *180*128.

(235) Warinner, C.; Hendy, J.; Speller, C.; Cappellini, E.; Fischer, R.; Trachsel, C.; Arneborg, J.; Lynnerup, N.; Craig, O. E.; Swallow, D.

M.; et al. Direct Evidence of Milk Consumption from Ancient Human Dental Calculus. *Sci. Rep.* **2014**, *4*, 1−6.

(236) Sotomayor-Vivas, C.; Hernández-Lemus, E.; Dorantes-Gilardi, R. Linking Protein Structural and Functional Change to Mutation Using Amino Acid Networks. *PLoS One* **2022**, *17*, No. e0261829.

(237) Guo, H. H.; Choe, J.; Loeb, L. A. Protein Tolerance to Random Amino Acid Change. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 9205−9210.

(238) Swire, J. Selection on Synthesis Cost Affects Interprotein Amino Acid Usage in All Three Domains of Life. *J. Mol. Evol.* **2007**, *64*, 558−571.

(239) Kumar, A.; Biswas, P. Effect of Site-Directed Point Mutations on Protein Misfolding: A Simulation Study. *Proteins* **2019**, *87*, 760−773.

(240) Zou, Z.; Zhang, J. Amino Acid Exchangeabilities Vary Across the Tree of Life. *Sci. Adv.* **2019**, *5*, No. eaax3124.

(241) Rokas, A.; Carroll, S. B. Frequent and Widespread Parallel Evolution of Protein Sequences. *Mol. Biol. Evol.* **2008**, *25*, 1943−1953.

(242) Horn, R. L.; Marques, A. J. D.; Manseau, M.; Golding, B.; Klütsch, C. F. C.; Abraham, K.; Wilson, P. J. Parallel Evolution of Site-Specific Changes in Divergent Caribou Lineages. *Ecol. Evol.* **2018**, *8*, 6053−6064.

(243) Lee, J.-H.; Lewis, K. M.; Moural, T. W.; Kirilenko, B.; Borgonovo, B.; Prange, G.; Koessl, M.; Huggenberger, S.; Kang, C.; Hiller, M. Molecular Parallelism in Fast-Twitch Muscle Proteins in Echolocating Mammals. *Sci. Adv.* **2018**, *4*, No. eaat9660.

(244) A von der Dunk, S. H.; Snel, B. Recurrent Sequence Evolution after Independent Gene Duplication. *BMC Evol. Biol.* **2020**, *20*, 1−17.

(245) Brandt, L. Ø.; Mannering, U. Taxonomic Identification of Danish Viking Age Shoes and Skin Objects by ZooMS (Zooarchaeology by Mass Spectrometry). *J. Proteomics* **2021**, *231*, 104038.

(246) Shoulders, M. D.; Raines, R. T. Collagen Structure and Stability. *Annu. Rev. Biochem.* **2009**, *78*, 929−958.

(247) Perumal, S.; Antipova, O.; Orgel, J. P. R. O. Collagen Fibril Architecture, Domain Organization, and Triple-Helical Conformation Govern Its Proteolysis. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 2824−2829.

(248) Henriksen, K.; Karsdal, M. A. Chapter 1 - Type I Collagen. In *Biochemistry of Collagens, Laminins and Elastin (Second ed.)*; Karsdal, M. A., Ed.; Academic Press, 2019; pp 1−12. DOI: 10.1016/B978-0-12-817068-7.00001-X.

(249) Gong, Y.; Li, L.; Gong, D.; Yin, H.; Zhang, J. Biomolecular Evidence of Silk from 8,500 Years Ago. *PLoS One* **2016**, *11*, No. e0168042.

(250) Sire, J.-Y.; Delgado, S.; Fromentin, D.; Girondot, M. Amelogenin: Lessons from Evolution. *Arch. Oral Biol.* **2005**, *50*, 205−212.

(251) Solazzo, C. Characterizing Historical Textiles and Clothing with Proteomics. *Conserv. patrim.* **2019**, *31*, 97−114.

(252) Coutu, A. N.; Whitelaw, G.; le Roux, P.; Sealy, J. Earliest Evidence for the Ivory Trade in Southern Africa: Isotopic and ZooMS Analysis of Seventh-Tenth Century Ad Ivory from KwaZulu-Natal. *Afr. Archaeol. Rev.* **2016**, *33*, 411−435.

(253) von Holstein, I. C. C.; Ashby, S. P.; van Doorn, N. L.; Sachs, S. M.; Buckley, M.; Meiri, M.; Barnes, I.; Brundle, A.; Collins, M. J. Searching for Scandinavians in Pre-Viking Scotland: Molecular Fingerprinting of Early Medieval Combs. *J. Archaeol. Sci.* **2014**, *41*, 1−6.

(254) Kirby, D. P.; Buckley, M.; Promise, E.; Trauger, S. A.; Holdcraft, T. R. Identification of Collagen-Based Materials in Cultural Heritage. *Analyst* **2013**, *138*, 4849−4858.

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

(256) Rick, T.; Harvey, V. L.; Buckley, M. Collagen Fingerprinting and the Chumash Billfish Fishery, Santa Barbara Channel, California, USA. *Archaeol. Anthropol. Sci.* **2019**, *11*, 6639−6648.

(257) Welker, F.; Soressi, M.; Rendu, W.; Hublin, J.-J.; Collins, M. Using ZooMS to Identify Fragmentary Bone from the Late Middle/Early Upper Palaeolithic Sequence of Les Cottés, France. *J. Archaeol. Sci.* **2015**, *54*, 279−286.

(258) Welker, F.; Hajdinjak, M.; Talamo, S.; Jaouen, K.; Dannemann, M.; David, F.; Julien, M.; Meyer, M.; Kelso, J.; Barnes, I.; et al. Palaeoproteomic Evidence Identifies Archaic Hominins Associated with the Châtelperronian at the Grotte Du Renne. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 11162−11167.

(259) Bradfield, J.; Forssman, T.; Spindler, L.; Antonites, A. R. Identifying the Animal Species Used to Manufacture Bone Arrowheads in South Africa. *Archaeol. Anthropol. Sci.* **2019**, *11*, 2419−2434.

(260) Desmond, A.; Barton, N.; Bouzouggar, A.; Douka, K.; Fernandez, P.; Humphrey, L.; Morales, J.; Turner, E.; Buckley, M. ZooMS Identification of Bone Tools from the North African Later Stone Age. *J. Archaeol. Sci.* **2018**, *98*, 149−157.

(261) Coutu, A. N.; Taurozzi, A. J.; Mackie, M.; Jensen, T. Z. T.; Collins, M. J.; Sealy, J. Palaeoproteomics Confirm Earliest Domesticated Sheep in Southern Africa Ca. 2000 BP. *Sci. Rep.* **2021**, *11*, 6631.

(262) Harvey, V. L.; Egerton, V. M.; Chamberlain, A. T.; Manning, P. L.; Sellers, W. I.; Buckley, M. Interpreting the Historical Terrestrial Vertebrate Biodiversity of Cayman Brac (Greater Antilles, Caribbean) through Collagen Fingerprinting. *Holocene* **2019**, *29*, 531−542.

(263) Buckley, M.; Gu, M.; Shameer, S.; Patel, S.; Chamberlain, A. T. High-Throughput Collagen Fingerprinting of Intact Microfaunal Remains; A Low-Cost Method for Distinguishing between Murine Rodent Bones. *Rapid Commun. Mass Spectrom.* **2016**, *30*, 805−812.

(264) Speller, C.; van den Hurk, Y.; Charpentier, A.; Rodrigues, A.; Gardeisen, A.; Wilkens, B.; McGrath, K.; Rowsell, K.; Spindler, L.; Collins, M.; et al. Barcoding the Largest Animals on Earth: Ongoing Challenges and Molecular Solutions in the Taxonomic Identification of Ancient Cetaceans. *Philos. Trans. R. Soc. London B Biol. Sci.* **2016**, *371* (1702), 20150332.

(265) Buckley, M.; Cosgrove, R.; Garvey, J.; Prideaux, G. J. Identifying Remains of Extinct Kangaroos in Late Pleistocene Deposits Using Collagen Fingerprinting. *J. Quat. Sci.* **2017**, *32*, 653−660.

(266) Buckley, M.; Cheylan, M. Collagen Fingerprinting for the Species Identification of Archaeological Amphibian Remains. *Boreas* **2020**, *5*, 1.

(267) van der Sluis, L. G.; Hollund, H. I.; Buckley, M.; De Louw, P. G. B.; Rijsdijk, K. F.; Kars, H. Combining Histology, Stable Isotope Analysis and ZooMS Collagen Fingerprinting to Investigate the Taphonomic History and Dietary Behaviour of Extinct Giant Tortoises from the Mare Aux Songes Deposit on Mauritius. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **2014**, *416*, 80−91.

(268) Martisius, N. L.; Welker, F.; Dogandžić, T.; Grote, M. N.; Rendu, W.; Sinet-Mathiot, V.; Wilcke, A.; McPherron, S. J. P.; Soressi, M.; Steele, T. E. Non-Destructive ZooMS Identification Reveals Strategic Bone Tool Raw Material Selection by Neandertals. *Sci. Rep.* **2020**, *10*, 7746.

(269) Pétillon, J.-M.; Chauvière, F.-X.; Speller, C.; McGrath, K.; Rodrigues, A. S. L.; Charpentier, A.; Baleux, F. A Gray Whale in Magdalenian Perigord. Species Identification of a Bone Projectile Point from La Madeleine (Dordogne, France) Using Collagen Fingerprinting. *Paĺeo* **2019**, *30−1*, 230−242.

(270) Moreau, L.; Draily, C.; Cordy, J.-M.; Boyle, K.; Buckley, M.; Gjesfjeld, E.; Filzmoser, P.; Borgia, V.; Gibson, S. A.; Day, J.; et al. Adaptive Trade-Offs Towards the Last Glacial Maximum in North-Western Europe: A Multidisciplinary View from Walou Cave. *J. Paleolit. Archaeol.* **2021**, *4*, 11.

(271) Ashby, S. P.; Coutu, A. N.; Sindbæk, S. M. Urban Networks and Arctic Outlands: Craft Specialists and Reindeer Antler in Viking Towns. *Eur. J. Archaeol.* **2015**, *18*, 679−704.

(272) Brandt, L. Ø.; Ebsen, J. A.; Haase, K. Leather Shoes in Early Danish Cities: Choices of Animal Resources and Specialization of Crafts in Viking and Medieval Denmark. *Eur. J. Archaeol.* **2020**, *23*, 428−450.

(273) Ebsen, J. A.; Haase, K.; Larsen, R.; Sommer, D. V. P.; Brandt, L. Ø. Identifying Archaeological Leather - Discussing the Potential of Grain Pattern Analysis and Zooarchaeology by Mass Spectrometry (ZooMS) through a Case Study Involving Medieval Shoe Parts from Denmark. *J. Cult. Herit.* **2019**, *39*, 21−31.

(274) Ives, J. W.; Froese, D.; Collins, M.; Brock, F. Radiocarbon and Protein Analyses Indicate an Early Holocene Age for the Osseous Rod from Grenfell, Saskatchewan, Canada. *Am. Antiq.* **2014**, *79*, 782−793.

(275) Sinet-Mathiot, V.; Smith, G. M.; Romandini, M.; Wilcke, A.; Peresani, M.; Hublin, J.-J.; Welker, F. Combining ZooMS and Zooarchaeology to Study Late Pleistocene Hominin Behaviour at Fumane (Italy). *Sci. Rep.* **2019**, *9*, 12350.

(276) Gron, K. J.; Rowley-Conwy, P.; Jensen, T. Z. T.; Taurozzi, A. J.; Marciniak, A. Separating Caprine (*Capra/Ovis*) Distal Tibiae: A Case Study from the Polish Neolithic. *Int. J. Osteoarchaeol.* **2020**, *30*, 170−179.

(277) Pilaar Birch, S. E.; Scheu, A.; Buckley, M.; Çakırlar, C. Combined Osteomorphological, Isotopic, aDNA, and ZooMS Analyses of Sheep and Goat Remains from Neolithic Ulucak, Turkey. *Archaeol. Anthropol. Sci.* **2019**, *11*, 1669−1681.

(278) Taylor, W.; Shnaider, S.; Abdykanova, A.; Fages, A.; Welker, F.; Irmer, F.; Seguin-Orlando, A.; Khan, N.; Douka, K.; et al. Early Pastoral Economies along the Ancient Silk Road: Biomolecular Evidence from the Alay Valley, Kyrgyzstan. *PLoS One* **2018**, *13*, No. e0205646.

(279) Culley, C.; Janzen, A.; Brown, S.; Prendergast, M. E.; Shipton, C.; Ndiema, E.; Petraglia, M. D.; Boivin, N.; Crowther, A. Iron Age Hunting and Herding in Coastal Eastern Africa: ZooMS Identification of Domesticates and Wild Bovids at Panga Ya Saidi, Kenya. *J. Archaeol. Sci.* **2021**, *130*, 105368.

(280) Prendergast, M. E.; Quintana Morales, E. M.; Crowther, A.; Horton, M. C.; Boivin, N. L. Dietary Diversity on the Swahili Coast: The Fauna from Two Zanzibar Trading Locales. *Int. J. Osteoarchaeol.* **2017**, *27*, 621−637.

(281) Prendergast, M. E.; Buckley, M.; Crowther, A.; Frantz, L.; Eager, H.; Lebrasseur, O.; Hutterer, R.; Hulme-Beaman, A.; Van Neer, W.; Douka, K.; et al. Reconstructing Asian Faunal Introductions to Eastern Africa from Multi-Proxy Biomolecular and Archaeological Datasets. *PLoS One* **2017**, *12*, No. e0182565.

(282) Guiry, E.; Buckley, M. Urban Rats Have Less Variable, Higher Protein Diets. *Proc. Biol. Sci.* **2018**, DOI: 10.1098/rspb.2018.1441.

(283) Oueslati, T.; Kbiri Alaoui, M.; Ichkhakh, A.; Callegarin, L.; de Chazelle, C.-A.; Rocca, E.; Carrato, et C. 1st Century BCE Occurrence of Chicken, House Mouse and Black Rat in Morocco: Socio-Economic Changes around the Reign of Juba II on the Site of Rirha. *J. Archaeol. Sci. Rep.* **2020**, *29*, 102162.

(284) Buckley, M.; Kansa, S. W. Collagen Fingerprinting of Archaeological Bone and Teeth Remains from Domuztepe, South Eastern Turkey. *Archaeol. Anthropol. Sci.* **2011**, *3*, 271−280.

(285) van den Hurk, Y.; Spindler, L.; McGrath, K.; Speller, C. Medieval Whalers in the Netherlands and Flanders: Zooarchaeological Analysis of Medieval Cetacean Remains. *Environ. Archaeol.* **2020**, 1−15.

(286) Evans, S.; Briz i Godino, I.; Álvarez, M.; Rowsell, K.; Collier, P.; de Goodall, R. N. P.; Mulville, J.; Lacrouts, A.; Collins, M. J.; Speller, C. Using Combined Biomolecular Methods to Explore Whale Exploitation and Social Aggregation in Hunter-Gatherer-Fisher Society in Tierra Del Fuego. *J. Archaeol. Sci. Rep.* **2016**, *6*, 757−767.

(287) Ma, J.; Wang, Y.; Baryshnikov, G. F.; Drucker, D. G.; McGrath, K.; Zhang, H.; Bocherens, H.; Hu, Y. The Mammuthus-Coelodonta Faunal Complex at Its Southeastern Limit: A Biogeochemical Paleoecology Investigation in Northeast Asia. *Quat. Int.* **2021**, *591*, 93−106.

(288) Winter, R. M.; de Kock, W.; Palsbøll, P. J.; Çakirlar, C. Potential Applications of Biomolecular Archaeology to the Ecohistory of Sea Turtles and Groupers in Levant Coastal Antiquity. *J. Archaeol. Sci. Rep.* **2021**, *36*, 102872.

(289) Garrison, E. G.; Morgan, G. S.; McGrath, K.; Speller, C.; Cherkinsky, A. Recent Dating of Extinct Atlantic Gray Whale Fossils, (*Eschrichtius Robustus*), Georgia Bight and Florida, Western Atlantic Ocean. *PeerJ.* **2019**, *7*, No. e6381.

(290) Devièse, T.; Karavanić, I.; Comeskey, D.; Kubiak, C.; Korlević, P.; Hajdinjak, M.; Radović, S.; Procopio, N.; Buckley, M.; Pääbo, S.; et al. Direct Dating of Neanderthal Remains from the Site of Vindija Cave and Implications for the Middle to Upper Paleolithic Transition. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 10606−10611.

(291) Slon, V.; Mafessoni, F.; Vernot, B.; de Filippo, C.; Grote, S.; Viola, B.; Hajdinjak, M.; Peyrégne, S.; Nagel, S.; Brown, S.; et al. The Genome of the Offspring of a Neanderthal Mother and a Denisovan Father. *Nature* **2018**, *561*, 113−116.

(292) Brown, S.; Higham, T.; Slon, V.; Pääbo, S.; Meyer, M.; Douka, K.; Brock, F.; Comeskey, D.; Procopio, N.; Shunkov, M.; et al. Identification of a New Hominin Bone from Denisova Cave, Siberia Using Collagen Fingerprinting and Mitochondrial DNA Analysis. *Sci. Rep.* **2016**, *6*, 23559.

(293) Fiddyment, S.; Teasdale, M. D.; Vnouček, J.; Lévêque, É.; Binois, A.; Collins, M. J. So You Want to Do Biocodicology? A Field Guide to the Biological Analysis of Parchment. *Herit. Sci.* **2019**, *7*, 1−10.

(294) Teasdale, M. D.; van Doorn, N. L.; Fiddyment, S.; Webb, C. C.; O'Connor, T.; Hofreiter, M.; Collins, M. J.; Bradley, D. G. Paging through History: Parchment as a Reservoir of Ancient DNA for Next Generation Sequencing. *Philos. Trans. R. Soc. London B Biol. Sci.* **2015**, *370*, 20130379.

(295) Teasdale, M. D.; Fiddyment, S.; Vnouček, J.; Mattiangeli, V.; Speller, C.; Binois, A.; Carver, M.; Dand, C.; Newfield, T. P.; Webb, C. C.; et al. The York Gospels: A 1000-Year Biological Palimpsest. *R. Soc. Open Sci.* **2017**, *4*, 170988.

(296) Ruffini-Ronzani, N.; Nieus, J.-F.; Soncin, S.; Hickinbotham, S.; Dieu, M.; Bouhy, J.; Charles, C.; Ruzzier, C.; Falmagne, T.; Hermand, X.; et al. A Biocodicological Analysis of the Medieval Library and Archive from Orval Abbey, Belgium. *R. Soc. Open Sci.* **2021**, *8*, 210210.

(297) Sommer, D. V. P.; Mühlen Axelsson, K.; Collins, M. J.; Fiddyment, S.; Bredal-Jørgensen, J.; Simonsen, K. P.; Lauridsen, C. B.; Larsen, R. Multiple Microanalyses of a Sample from the Vinland Map. *Archaeometry* **2017**, *59*, 287−301.

(298) Vnouček, J.; Fiddyment, S.; Quandt, A.; Rabitsch, S.; Collins, M.; Hofmann, C. The Parchment of the Vienna Genesis: Characteristics and Manufacture. In *The Vienna Genesis Material Analysis and Conservation of a Late Antique Illuminated Manuscript on Purple Parchment*; Boehlau Verlag GmbH & Co. KG, 2020; pp 35−70.

(299) Doherty, S. P.; Henderson, S.; Fiddyment, S.; Finch, J.; Collins, M. J. Scratching the Surface: The Use of Sheepskin Parchment to Deter Textual Erasure in Early Modern Legal Deeds. *Herit. Sci.* **2021**, *9*, 1−6.

(300) Calà, E.; Agostino, A.; Fenoglio, G.; Capra, V.; Porticelli, F.; Manzari, F.; Fiddyment, S.; Aceto, M. The Messale Rosselli: Scientific Investigation on an Outstanding 14th Century Illuminated Manuscript from Avignon. *J. Archaeol. Sci. Rep.* **2019**, *23*, 721−730.

(301) Fiddyment, S.; Collins, M. From Field to Frame. The Contribution of Bioarchaeological Methods to Understanding Parchment Production. *Gaz. du livre médiév.* **2017**, *63*, 55−63.

(302) Gordon, J. In the Flesh? Anthropodermic Bibliopegy Verification and Its Implications. *RBM: A Journal of Rare Books, Manuscripts, and Cultural Heritage* **2017**, DOI: 10.5860/rbm.17.2.9664.

(303) 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* **2012**, *75*, 3365−3373.

(304) McKittrick, J.; Chen, P.-Y.; Bodde, S. G.; Yang, W.; Novitskaya, E. E.; Meyers, M. A. The Structure, Functions, and Mechanical Properties of Keratin. *JOM* **2012**, *64*, 449−468.

(305) Smack, D. P.; Korge, B. P.; James, W. D. Keratin and Keratinization. *J. Am. Acad. Dermatol.* **1994**, *30*, 85−102.

(306) Marshall, R. C.; Orwin, D. F.; Gillespie, J. M. Structure and Biochemistry of Mammalian Hard Keratin. *Electron Microsc. Rev.* **1991**, *4*, 47−83.

(307) Dhouailly, D.; Godefroit, P.; Martin, T.; Nonchev, S.; Caraguel, F.; Oftedal, O. Getting to the Root of Scales, Feather and Hair: As Deep as Odontodes? *Exp. Dermatol.* **2019**, *28*, 503−508.

(308) Eckhart, L.; Ehrlich, F. Evolution of Trichocyte Keratins. *Adv. Exp. Med. Biol.* **2018**, *1054*, 33−45.

(309) Solazzo, C.; Wadsley, M.; Dyer, J. M.; Clerens, S.; Collins, M. J.; Plowman, J. Characterisation of Novel α-keratin Peptide Markers for Species Identification in Keratinous Tissues Using Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2013**, *27*, 2685−2698.

(310) O'Connor, S.; Solazzo, C.; Collins, M. Advances in Identifying Archaeological Traces of Horn and Other Keratinous Hard Tissues. *Stud. Conserv.* **2015**, *60*, 393−417.

(311) Wegst, U. G. K.; Ashby, M. F. The Mechanical Efficiency of Natural Materials. *Philos. Mag.* **2004**, *84*, 2167−2186.

(312) Solazzo, C.; Heald, S.; Ballard, M. W.; Ashford, D. A.; DePriest, P. T.; Koestler, R. J.; Collins, M. J. Proteomics and Coast Salish Blankets: A Tale of Shaggy Dogs? *Antiquity* **2011**, *85*, 1418−1432.

(313) Azémard, C.; Zazzo, A.; Marie, A.; Lepetz, S.; Debaine-Francfort, C.; Idriss, A.; Zirah, S. Animal Fibre Use in the Keriya Valley (Xinjiang, China) during the Bronze and Iron Ages: A Proteomic Approach. *J. Archaeol. Sci.* **2019**, *110*, 104996.

(314) Solazzo, C. Follow-up on the Characterization of Peptidic Markers in Hair and Fur for the Identification of Common North American Species. *Rapid Commun. Mass Spectrom.* **2017**, *31*, 1375−1384.

(315) Azémard, C.; Dufour, E.; Zazzo, A.; Wheeler, J. C.; Goepfert, N.; Marie, A.; Zirah, S. Untangling the Fibre Ball: Proteomic Characterization of South American Camelid Hair Fibres by Untargeted Multivariate Analysis and Molecular Networking. *J. Proteomics* **2021**, *231*, 104040.

(316) Gong, H.; Zhou, H.; McKenzie, G. W.; Hickford, J. G.; Yu, Z.; Clerens, S.; Dyer, J. M.; Plowman, J. E. Emerging Issues with the Current Keratin-Associated Protein Nomenclature. *Int. J. Trichology* **2010**, *2*, 104−105.

(317) Hollemeyer, K.; Altmeyer, W.; Heinzle, E.; Pitra, C. Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry Combined with Multidimensional Scaling, Binary Hierarchical Cluster Tree and Selected Diagnostic Masses Improves Species Identification of Neolithic Keratin Sequences from Furs of the Tyrolean Iceman Oetzi. *Rapid Commun. Mass Spectrom.* **2012**, *26*, 1735−1745.

(318) Solazzo, C.; Fitzhugh, W.; Kaplan, S.; Potter, C.; Dyer, J. M. Molecular Markers in Keratins from Mysticeti Whales for Species Identification of Baleen in Museum and Archaeological Collections. *PLoS One* **2017**, *12*, No. e0183053.

(319) Solazzo, C.; Soulat, J.; Cleland, T. Creation of a Peptide Database of Corneous Beta-Proteins of Marine Turtles for the Identification of Tortoiseshell: Archaeological Combs as Case Study. *R Soc. Open Sci.* **2021**, *8*, 201857.

(320) Wang, B.; Gu, J.; Chen, B.; Xu, C.; Zheng, H.; Peng, Z.; Zhou, Y.; Hu, Z. Development of an Enzyme-Linked Immunosorbent Assay and Gold-Labelled Immunochromatographic Strip Assay for the Detection of Ancient Wool. *J. Anal. Methods Chem.* **2018**, *2641624*.

(321) Wang, B.; Gu, J.; You, Q.; Chen, B.; Zheng, H.; Zhou, Y.; Hu, Z. Preparation of Artificial Antibodies and Development of an Antibody-Based Indirect ELISA for the Detection of Ancient Wool. *Anal. Methods* **2018**, *10*, 1480−1487.

(322) Plowman, J.; Thomas, A.; Perloiro, T.; Clerens, S.; de Almeida, A. M. Characterisation of White and Black Merino Wools: A Proteomics Study. *Animal* **2019**, *13*, 659−665.

(323) Plowman, J. E.; Harland, D. P.; Campos, A. M. O.; Rocha E Silva, S.; Thomas, A.; Vernon, J. A.; van Koten, C.; Hefer, C.; Clerens, S.; de Almeida, A. M. The Wool Proteome and Fibre Characteristics of Three Distinct Genetic Ovine Breeds from Portugal. *J. Proteomics* **2020**, *225*, 103853.

(324) Chu, F.; Mason, K. E.; Anex, D. S.; Jones, A. D.; Hart, B. R. Hair Proteome Variation at Different Body Locations on Genetically

(325) Variant Peptide Detection for Protein-Based Human Identification. *Sci. Rep.* **2019**, *9*, 7641.

(325) Adav, S. S.; Subbaiaih, R. S.; Kerk, S. K.; Lee, A. Y.; Lai, H. Y.; Ng, K. W.; Sze, S. K.; Schmidtchen, A. Studies on the Proteome of Human Hair - Identification of Histones and Deamidated Keratins. *Sci. Rep.* **2018**, *8*, 1599.

(326) Guo, C. 6 - Insect and Animal-Originated Fibres: Silk and Wool. In *Fundamentals of Natural Fibres and Textiles*; Mondal, M. I. H., Ed.; Woodhead Publishing, 2021; pp 153−178 DOI: 10.1016/B978-0-12-821483-1.00020-6.

(327) Omenetto, F. G.; Kaplan, D. L. New Opportunities for an Ancient Material. *Science* **2010**, *329*, 528−531.

(328) Good, I. L.; Kenoyer, J. M.; Meadow, R. H. New Evidence for Early Silk in the Indus Civilization. *Archaeometry* **2009**, *51*, 457−466.

(329) Good, I. The Archaeology of Early Silk. *Silk Roads, Other Roads: Textile Society of America 8th Biennial Symposium.* **2002**.

(330) Holland, C.; Numata, K.; Rnjak-Kovacina, J.; Seib, F. P. The Biomedical Use of Silk: Past, Present, Future. *Adv. Healthc. Mater.* **2019**, *8*, No. e1800465.

(331) Chen, R.; Zhu, C.; Hu, M.; Zhou, L.; Yang, H.; Zheng, H.; Zhou, Y.; Hu, Z.; Peng, Z.; Wang, B. Comparative Analysis of Proteins from Bombyx Mori and Antheraea Pernyi Cocoons for the Purpose of Silk Identification. *J. Proteomics* **2019**, *209*, 103510.

(332) Li, L.; Zhu, L.; Xie, Y. Proteomics Analysis of the Soil Textile Imprints from Tomb M6043 of the Dahekou Cemetery Site in Yicheng County, Shanxi Province, China. *Archaeol. Anthropol. Sci.* **2021**, *13*. DOI: 10.1007/s12520-020-01258-0.

(333) Yu, X.; Gong, Y.; Tan, P.; Shu, J.; Liu, F.; Yang, J.; Gong, D. Mass Spectrometry Analysis of Textile Used in Decorating the Coronet of Empress Xiao of the Sui Dynasty (581−618 A.D.). *J. Cult. Herit.* **2017**, *25*, 185−188.

(334) Zheng, H.; Yang, H.; Zhang, W.; Yang, R.; Su, B.; Zhao, X.; Zhou, Y.; Dai, X. Insight of Silk Relics of Mineralized Preservation in Maoling Mausoleum Using Two Enzyme-Linked Immunological Methods. *J. Archaeol. Sci.* **2020**, *115*, 105089.

(335) Sasaki, S.; Shimokawa, H. The Amelogenin Gene. *Int. J. Dev. Biol.* **1995**, *39*, 127−133.

(336) Lau, E. C.; Mohandas, T. K.; Shapiro, L. J.; Slavkin, H. C.; Snead, M. L. Human and Mouse Amelogenin Gene Loci Are on the Sex Chromosomes. *Genomics* **1989**, *4*, 162−168.

(337) Delgado, S.; Girondot, M.; Sire, J.-Y. Molecular Evolution of Amelogenin in Mammals. *J. Mol. Evol.* **2005**, *60*, 12−30.

(338) Mannucci, A.; Sullivan, K. M.; Ivanov, P. L.; Gill, P. Forensic Application of a Rapid and Quantitative DNA Sex Test by Amplification of the X-Y Homologous Gene Amelogenin. *Int. J. Legal Med.* **1994**, *106*, 190−193.

(339) Nagare, S. P.; Chaudhari, R. S.; Birangane, R. S.; Parkarwar, P. C. Sex Determination in Forensic Identification, a Review. *J. Forensic Dent. Sci.* **2018**, *10*, 61−66.

(340) Ennis, S.; Gallagher, T. F. A PCR-Based Sex-Determination Assay in Cattle Based on the Bovine Amelogenin Locus. *Anim. Genet.* **1994**, *25*, 425−427.

(341) Weikard, R.; Pitra, C.; Kühn, C. Amelogenin Cross-Amplification in the Family Bovidae and Its Application for Sex Determination. *Mol. Reprod. Dev.* **2006**, *73*, 1333−1337.

(342) Pfeiffer, I.; Brenig, B. X- and Y-Chromosome Specific Variants of the Amelogenin Gene Allow Sex Determination in Sheep (Ovis Aries) and European Red Deer (Cervus Elaphus). *BMC Genet.* **2005**, *6*, 16.

(343) Tsai, T. C.; Wu, S. H.; Chen, H. L.; Tung, Y. T.; Cheng, W. T. K.; Huang, J. C.; Chen, C. M. Identification of Sex-Specific Polymorphic Sequences in the Goat Amelogenin Gene for Embryo Sexing. *J. Anim. Sci.* **2011**, *89*, 2407−2414.

(344) Fontanesi, L.; Scotti, E.; Russo, V. Differences of the Porcine Amelogenin X and Y Chromosome Genes (AMELX and AMELY) and Their Application for Sex Determination in Pigs. *Mol. Reprod. Dev.* **2008**, *75*, 1662−1668.

(345) Hasegawa, T.; Sato, F.; Ishida, N.; Fukushima, Y.; Mukoyama, H. Sex Determination by Simultaneous Amplification of Equine SRY and Amelogenin Genes. *J. Vet. Med. Sci.* **2000**, *62*, 1109−1110.

(346) Yamamoto, K.; Tsubota, T.; Komatsu, T.; Katayama, A.; Murase, T.; Kita, I.; Kudo, T. Sex Identification of Japanese Black Bear, Ursus Thibetanus Japonicus, by PCR Based on Amelogenin Gene. *J. Vet. Med. Sci.* **2002**, *64*, 505−508.

(347) Nielsen-Marsh, C. M.; Stegemann, C.; Hoffmann, R.; Smith, T.; Feeney, R.; Toussaint, M.; Harvati, K.; Panagopoulou, E.; Hublin, J.-J.; Richards, M. P. Extraction and Sequencing of Human and Neanderthal Mature Enamel Proteins Using MALDI-TOF/TOF MS. *J. Archaeol. Sci.* **2009**, *36*, 1758−1763.

(348) Porto, I. M.; Laure, H. J.; Tykot, R. H.; de Sousa, F. B.; Rosa, J. C.; Gerlach, R. F. Recovery and Identification of Mature Enamel Proteins in Ancient Teeth. *Eur. J. Oral Sci.* **2011**, *119* (Suppl 1), 83−87.

(349) Parker, G. J.; Yip, J. M.; Eerkens, J. W.; Salemi, M. Sex Estimation Using Sexually Dimorphic Amelogenin Protein Fragments in Human Enamel. *Journal of Archaeological Science* **2019**, *101*, 169−180.

(350) Lugli, F.; Di Rocco, G.; Vazzana, A.; Genovese, F.; Pinetti, D.; Cilli, E.; Carile, M. C.; Silvestrini, S.; Gabanini, G.; Arrighi, S.; et al. Enamel Peptides Reveal the Sex of the Late Antique "Lovers of Modena. *Sci. Rep.* **2019**, *9*, 13130.

(351) Lugli, F.; Figus, C.; Silvestrini, S.; Costa, V.; Bortolini, E.; Conti, S.; Peripoli, B.; Nava, A.; Sperduti, A.; Lamanna, L.; et al. Sex-Related Morbidity and Mortality in Non-Adult Individuals from the Early Medieval Site of Valdaro (Italy): The Contribution of Dental Enamel Peptide Analysis. *Journal of Archaeological Science: Reports* **2020**, *34*, 102625.

(352) Haas, R.; Watson, J.; Buonasera, T.; Southon, J.; Chen, J. C.; Noe, S.; Smith, K.; Llave, C. V.; Eerkens, J.; Parker, G. Female Hunters of the Early Americas. *Sci. Adv.* **2020**, DOI: 10.1126/sciadv.abd0310.

(353) Rebay-Salisbury, K.; Janker, L.; Pany-Kucera, D.; Schuster, D.; Spannagl-Steiner, M.; Waltenberger, L.; Salisbury, R. B.; Kanz, F. Child Murder in the Early Bronze Age: Proteomic Sex Identification of a Cold Case from Schleinbach, Austria. *Archaeol. Anthropol. Sci.* **2020**, *12*, 265.

(354) Buonasera, T.; Eerkens, J.; de Flamingh, A.; Engbring, L.; Yip, J.; Li, H.; Haas, R.; DiGiuseppe, D.; Grant, D.; Salemi, M.; et al. A Comparison of Proteomic, Genomic, and Osteological Methods of Archaeological Sex Estimation. *Sci. Rep.* **2020**, *10*, 11897.

(355) Nogueira, F. C. S.; Neves, L. X.; Pessoa-Lima, C.; Langer, M. C.; Domont, G. B.; Line, S. R. P.; Paes Leme, A. F.; Gerlach, R. F. Ancient Enamel Peptides Recovered from the South American Pleistocene Species Notiomastodon Platensis and Myocastor Cf. Coypus. *J. Proteomics* **2021**, *240*, 104187.

(356) Stamfelj, I. Sex Estimation Based on the Analysis of Enamel Peptides: False Assignments due to AMELY Deletion. *J. Archaeol. Sci.* **2021**, *130*, 105345.

(357) Parker, G. J.; Buonasera, T.; Yip, J. M.; Eerkens, J. W.; Salemi, M.; Durbin-Johnson, B.; Haas, R.; Buikstra, J. E.; Klaus, H.; Rocke, D. M.; Phinney, B. AMELY Deletion Is Not Detected in Systematically Sampled Reference Populations: A Reply to Štamfelj. *J. Archaeol. Sci.* **2021**, *130*, 105354.

(358) Wu, L.; Han, D. K. Overcoming the Dynamic Range Problem in Mass Spectrometry-Based Shotgun Proteomics. *Expert Rev. Proteomics* **2006**, *3*, 611−619.

(359) Uhlén, M.; Fagerberg, L.; Hallström, B. M.; Lindskog, C.; Oksvold, P.; Mardinoglu, A.; Sivertsson, Å.; Kampf, C.; Sjöstedt, E.; Asplund, A.; et al. Proteomics. Tissue-Based Map of the Human Proteome. *Science* **2015**, *347*, 1260419.

(360) Salmon, C. R.; Giorgetti, A. P. O.; Paes Leme, A. F.; Domingues, R. R.; Kolli, T. N.; Foster, B. L.; Nociti, F. H., Jr. Microproteome of Dentoalveolar Tissues. *Bone* **2017**, *101*, 219−229.

(361) Boskey, A. L.; Robey, P. G. The Composition of Bone. In *Primer on the Metabolic Bone Diseases and Disorders of Mineral Metabolism*; Wiley, 2008; pp 84−92 DOI: 10.1002/9781119266594.ch11.

(362) Lin, X.; Patil, S.; Gao, Y.-G.; Qian, A. The Bone Extracellular Matrix in Bone Formation and Regeneration. *Front. Pharmacol.* **2020**, DOI: 10.3389/fphar.2020.00757.

(363) Widbiller, M.; Schweikl, H.; Bruckmann, A.; Rosendahl, A.; Hochmuth, E.; Lindner, S. R.; Buchalla, W.; Galler, K. M. Shotgun Proteomics of Human Dentin with Different Prefractionation Methods. *Sci. Rep.* **2019**, *9*, 4457.

(364) Sawafuji, R.; Cappellini, E.; Nagaoka, T.; Fotakis, A. K.; Jersie-Christensen, R. R.; Olsen, J. V.; Hirata, K.; Ueda, S. Proteomic Profiling of Archaeological Human Bone. *R Soc. Open Sci.* **2017**, *4*, 161004.

(365) 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. A Late Middle Pleistocene Denisovan Mandible from the Tibetan Plateau. *Nature* **2019**, *569*, 409−412.

(366) Kostyukevich, Y.; Bugrova, A.; Chagovets, V.; Brzhozovskiy, A.; Indeykina, M.; Vanyushkina, A.; Zherebker, A.; Mitina, A.; Kononikhin, A.; Popov, I.; et al. Proteomic and Lipidomic Analysis of Mammoth Bone by High-Resolution Tandem Mass Spectrometry Coupled with Liquid Chromatography. *Eur. J. Mass Spectrom.* **2018**, *24*, 411−419.

(367) Barker, A.; Dombrosky, J.; Chaput, D.; Venables, B.; Wolverton, S.; Stevens, S. M. Validation of a Non-Targeted LC−MS Approach for Identifying Ancient Proteins: Method Development on Bone to Improve Artifact Residue Analysis. *Ethnobiol. Lett.* **2015**, *6*, 162−174.

(368) Rüther, P. L.; Husic, I. M.; Bangsgaard, P.; Murphy Gregersen, K.; Pantmann, P.; Carvalho, M.; Godinho, R. M.; Friedl, L.; Cascalheira, J.; Jørkov, M. L. S. SPIN - Species by Proteome INvestigation. *BioRxiv* **2021**, DOI: 10.1101/2021.02.23.432520.

(369) Doherty, A. H.; Roteliuk, D. M.; Gookin, S. E.; McGrew, A. K.; Broccardo, C. J.; Condon, K. W.; Prenni, J. E.; Wojda, S. J.; Florant, G. L.; Donahue, S. W. Exploring the Bone Proteome to Help Explain Altered Bone Remodeling and Preservation of Bone Architecture and Strength in Hibernating Marmots. *Physiol. Biochem. Zool.* **2016**, *89*, 364−376.

(370) Taha, I. N.; Naba, A. Exploring the Extracellular Matrix in Health and Disease Using Proteomics. *Essays Biochem.* **2019**, *63*, 417−432.

(371) Johnson, A. A.; Shokhirev, M. N.; Wyss-Coray, T.; Lehallier, B. Systematic Review and Analysis of Human Proteomics Aging Studies Unveils a Novel Proteomic Aging Clock and Identifies Key Processes That Change with Age. *Ageing Res. Rev.* **2020**, *60*, 101070.

(372) Masters, E. A.; Trombetta, R. P.; de Mesy Bentley, K. L.; Boyce, B. F.; Gill, A. L.; Gill, S. R.; Nishitani, K.; Ishikawa, M.; Morita, Y.; Ito, H.; et al. Evolving Concepts in Bone Infection: Redefining "Biofilm", "Acute vs. Chronic Osteomyelitis", "the Immune Proteome" and "Local Antibiotic Therapy. *Bone Res.* **2019**, *7*, 1−18.

(373) Olson, M. E.; Horswill, A. R. *Staphylococcus Aureus* Osteomyelitis: Bad to the Bone. *Cell Host Microbe.* **2013**, *13*, 629−631.

(374) Niforou, K. M.; Anagnostopoulos, A. K.; Vougas, K.; Kittas, C.; Gorgoulis, V. G.; Tsangaris, G. T. The Proteome Profile of the Human Osteosarcoma U2OS Cell Line. *Cancer Genomics Proteomics* **2008**, *5*, 63−78.

(375) Iglesias-Gato, D.; Thysell, E.; Tyanova, S.; Crnalic, S.; Santos, A.; Lima, T. S.; Geiger, T.; Cox, J.; Widmark, A.; Bergh, A.; Mann, M.; Flores-Morales, A.; Wikström, P. The Proteome of Prostate Cancer Bone Metastasis Reveals Heterogeneity with Prognostic Implications. *Clin. Cancer Res.* **2018**, *24*, 5433−5444.

(376) Schultz, M.; Parzinger, H.; Posdnjakov, D. V.; Chikisheva, T. A.; Schmidt-Schultz, T. H. Oldest Known Case of Metastasizing Prostate Carcinoma Diagnosed in the Skeleton of a 2,700-Year-Old Scythian King from Arzhan (Siberia, Russia). *Int. J. Cancer* **2007**, *121*, 2591−2595.

(377) Solovev, I.; Shaposhnikov, M.; Moskalev, A. Multi-Omics Approaches to Human Biological Age Estimation. *Mech. Ageing Dev.* **2020**, *185*, 111192.

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

(379) Johnston, E.; Buckley, M. Relative Protein Abundances and Biological Ageing in Whole Skeletal Elements. *J. Proteome Res.* **2021**, *20*, 538−548.

(380) Zapico, S. C.; DeGaglia, C. M.; Adserias-Garriga, J. Age Estimation Based on Chemical Approaches. In *Age Estimation*; Elsevier, 2019; pp 199−211 DOI: 10.1016/B978-0-12-814491-6.00014-5.

(381) Prieto-Bonete, G.; Pérez-Cárceles, M. D.; Maurandi-López, A.; Pérez-Martínez, C.; Luna, A. Association between Protein Profile and Postmortem Interval in Human Bone Remains. *J. Proteomics* **2019**, *192*, 54−63.

(382) Mizukami, H.; Hathway, B.; Procopio, N. Aquatic Decomposition of Mammalian Corpses: A Forensic Proteomic Approach. *J. Proteome Res.* **2020**, *19*, 2122−2135.

(383) Bell, P. A.; Solis, N.; Kizhakkedathu, J. N.; Matthew, I.; Overall, C. M. Proteomic and N-Terminomic TAILS Analyses of Human Alveolar Bone Proteins: Improved Protein Extraction Methodology and LysargiNase Digestion Strategies Increase Proteome Coverage and Missing Protein Identification. *J. Proteome Res.* **2019**, *18*, 4167−4179.

(384) Bai, Y.; Yu, Z.; Ackerman, L.; Zhang, Y.; Bonde, J.; Li, W.; Cheng, Y.; Habelitz, S. Protein Nanoribbons Template Enamel Mineralization. *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117*, 19201−19208.

(385) Margolis, H. C.; Beniash, E.; Fowler, C. E. Role of Macromolecular Assembly of Enamel Matrix Proteins in Enamel Formation. *J. Dent. Res.* **2006**, *85*, 775−793.

(386) Duverger, O.; Beniash, E.; Morasso, M. I. Keratins as Components of the Enamel Organic Matrix. *Matrix Biol.* **2016**, *52*, 260−265.

(387) Duverger, O.; Ohara, T.; Shaffer, J. R.; Donahue, D.; Zerfas, P.; Dullnig, A.; Crecelius, C.; Beniash, E.; Marazita, M. L.; Morasso, M. I. Hair Keratin Mutations in Tooth Enamel Increase Dental Decay Risk. *J. Clin. Invest.* **2014**, *124*, 5219−5224.

(388) Castiblanco, G. A.; Rutishauser, D.; Ilag, L. L.; Martignon, S.; Castellanos, J. E.; Mejía, W. Identification of Proteins from Human Permanent Erupted Enamel. *Eur. J. Oral Sci.* **2015**, *123*, 390−395.

(389) Welker, F. Palaeoproteomics for Human Evolution Studies. *Quat. Sci. Rev.* **2018**, *190*, 137−147.

(390) Marie, P.; Labas, V.; Brionne, A.; Harichaux, G.; Hennequet-Antier, C.; Nys, Y.; Gautron, J. Quantitative Proteomics and Bioinformatic Analysis Provide New Insight into Protein Function During Avian Eggshell Biomineralization. *J. Proteomics* **2015**, *113*, 178−193.

(391) Hincke, M. T.; Nys, Y.; Gautron, J. The Role of Matrix Proteins in Eggshell Formation. *J. Poult. Sci.* **2010**, *47*, 208−219.

(392) Le Roy, N.; Stapane, L.; Gautron, J.; Hincke, M. T. Evolution of the Avian Eggshell Biomineralization Protein Toolkit - New Insights from Multi-Omics. *Front. Genet.* **2021**, *12*, 672433.

(393) Yang, R.; Geng, F.; Huang, X.; Qiu, N.; Li, S.; Teng, H.; Chen, L.; Song, H.; Huang, Q. Integrated Proteomic, Phosphoproteomic and N-Glycoproteomic Analyses of Chicken Eggshell Matrix. *Food Chem.* **2020**, *330*, 127167.

(394) Fothergill, B.; Best, J.; Foster, A.; Demarchi, B. Hens, Health and Husbandry: Integrated Approaches to Past Poultry-Keeping in England. *Open Quat.* **2017**, *3*, 5.

(395) Stewart, J. R. M.; Allen, R. B.; Jones, A. K. G.; Kendall, T.; Penkman, K. E. H.; Demarchi, B.; O'Connor, T.; Collins, M. J. Walking on Eggshells: A Study of Egg Use in Anglo-Scandinavian York Based on Eggshell Identification Using ZooMS. *Int. J. Osteoarchaeol.* **2014**, *24*, 247−255.

(396) Maltby, M.; Allen, M.; Best, J.; Fothergill, B. T.; Demarchi, B. Counting Roman Chickens: Multidisciplinary Approaches to Human-Chicken Interactions in Roman Britain. *J. Archaeol. Sci. Rep.* **2018**, *19*, 1003−1015.

(397) Jonuks, T.; Oras, E.; Best, J.; Demarchi, B.; Mänd, R.; Presslee, S.; Vahur, S. Multi-Method Analysis of Avian Eggs as Grave Goods: Revealing Symbolism in Conversion Period Burials at Kukruse, NE Estonia. *Environ. Archaeol.* **2018**, *23*, 109−122.

(398) Demarchi, B.; Presslee, S.; Sakalauskaite, J.; Fischer, R.; Best, J. The Role of Birds at Çatalhöyük Revealed by the Analysis of Eggshell. *Quat. Int.* **2020**, *543*, 50−60.

(399) Demarchi, B.; Presslee, S.; Gutiérrez-Zugasti, I.; González-Morales, M.; Marín-Arroyo, A. B.; Straus, L. G.; Fischer, R. Birds of Prey and Humans in Prehistoric Europe: A View from El Mirón Cave, Cantabria (Spain). *J. Archaeol. Sci. Rep.* **2019**, *24*, 244−252.

(400) Kang, C. W.; Nam, K. T.; Olson, O. E.; Carlson, C. W. Effects of Dietary Protein Level, Restricted Feeding, Strain and Age on Eggshell Quality in Laying Jens. *Asian-australas. J. Anim. Sci.* **1996**, *9*, 727−735.

(401) Beacham, E. B.; Durand, S. R. Eggshell and the Archaeological Record: New Insights into Turkey Husbandry in the American Southwest. *J. Archaeol. Sci.* **2007**, *34*, 1610−1621.

(402) George, R. J.; Plog, S.; Watson, A. S.; Schmidt, K. L.; Culleton, B. J.; Harper, T. K.; Gilman, P. A.; LeBlanc, S. A.; Amato, G.; Whiteley, P.; et al. Archaeogenomic Evidence from the Southwestern US Points to a Pre-Hispanic Scarlet Macaw Breeding Colony. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, 8740−8745.

(403) Minnis, P. E.; Whalen, M. E.; Kelley, J. H.; Stewart, J. D. Prehistoric Macaw Breeding in the North American Southwest. *Am. Antiq.* **1993**, *58*, 270−276.

(404) Kocot, K. M.; Aguilera, F.; McDougall, C.; Jackson, D. J.; Degnan, B. M. Sea Shell Diversity and Rapidly Evolving Secretomes: Insights into the Evolution of Biomineralization. *Front. Zool.* **2016**, *13*, 23.

(405) Marin, F.; Marie, B.; Hamada, S. B.; Ramos-Silva, P.; Le Roy, N.; Guichard, N.; Wolf, S. E.; Montagnani, C.; Joubert, C.; Piquemal, D.; et al. Shellome': Proteins Involved in Mollusk Shell Biomineralization-Diversity, Functions. *Recent advances in pearl research* **2013**, *149*, 168.

(406) Sykes, G. A.; Collins, M. J.; Walton, D. I. The Significance of a Geochemically Isolated Intracrystalline Organic Fraction within Biominerals. *Org. Geochem.* **1995**, *23* (11), 1059−1065.

(407) 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.* **2008**, *3*, 2−25.

(408) Demarchi, B.; Rogers, K.; Fa, D. A.; Finlayson, C. J.; Milner, N.; Penkman, K. E. H. Intra-Crystalline Protein Diagenesis (IcPD) in Patella Vulgata. Part I: Isolation and Testing of the Closed System. *Quat. Geochronol.* **2013**, *16*, 144−157.

(409) Ortiz, J. E.; Sánchez-Palencia, Y.; Gutiérrez-Zugasti, I.; Torres, T.; González-Morales, M. Protein Diagenesis in Archaeological Gastropod Shells and the Suitability of This Material for Amino Acid Racemisation Dating: Phorcus Lineatus (da Costa, 1778). *Quat. Geochronol.* **2018**, *46*, 16−27.

(410) Jackson, D. J.; McDougall, C.; Woodcroft, B.; Moase, P.; Rose, R. A.; Kube, M.; Reinhardt, R.; Rokhsar, D. S.; Montagnani, C.; Joubert, C.; et al. Parallel Evolution of Nacre Building Gene Sets in Molluscs. *Mol. Biol. Evol.* **2010**, *27*, 591−608.

(411) Marin, F. Mollusc Shellomes: Past, Present and Future. *J. Struct. Biol.* **2020**, *212*, 107583.

(412) Sakalauskaite, J.; Plasseraud, L.; Thomas, J.; Albéric, M.; Thoury, M.; Perrin, J.; Jamme, F.; Broussard, C.; Demarchi, B.; Marin, F. The Shell Matrix of the European Thorny Oyster, Spondylus Gaederopus: Microstructural and Molecular Characterization. *J. Struct. Biol.* **2020**, *211*, 107497.

(413) Hoffmann, D. L.; Angelucci, D. E.; Villaverde, V.; Zapata, J.; Zilhão, J. Symbolic Use of Marine Shells and Mineral Pigments by Iberian Neandertals 115,000 Years Ago. *Sci. Adv.* **2018**, *4*, No. eaar5255.

(414) Lynnerup, N. Mummies. *Am. J. Phys. Anthropol.* **2007**, *134*, 162−190.

(415) Parkinson, E.; Skipp, P.; Aleksic, M.; Garrow, A.; Dadd, T.; Hughes, M.; Clough, G.; O'Connor, C. D. Proteomic Analysis of the Human Skin Proteome after *in Vivo* Treatment with Sodium Dodecyl Sulphate. *PLoS One* 2014, 9, e97772.

(416) Graham, H. K.; Eckersley, A.; Ozols, M.; Mellody, K. T.; Sherratt, M. J. Human Skin: Composition, Structure and Visualisation Methods. In *Skin Biophysics: From Experimental Characterisation to Advanced Modelling*; Limbert, G., Ed.; Springer International Publishing: Cham, 2019; pp 1–18 DOI: 10.1007/978-3-030-13279-8_1.

(417) Jones, J.; Mirzaei, M.; Ravishankar, P.; Xavier, D.; Lim, D. S.; Shin, D. H.; Bianucci, R.; Haynes, P. A. 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.* 2016, 374, 20150373.

(418) Corthals, A.; Koller, A.; Martin, D. W.; Rieger, R.; Chen, E. I.; Bernaski, M.; Recagno, G.; Dávalos, L. M. Detecting the Immune System Response of a 500 Year-Old Inca Mummy. *PLoS One* 2012, 7, No. e41244.

(419) Demarchi, B.; Boano, R.; Ceron, A.; Bello, F. D.; Favero-Longo, S. E.; Fiddyment, S.; Marochetti, E. F.; Mangiapane, G.; Mattonai, M.; Pennacini, C.; et al. Never Boring: Non-Invasive Palaeoproteomics of Mummified Human Skin. *J. Archaeol. Sci.* 2020, 119, 105145.

(420) Grimm, R. How Modern Mass Spectrometry Can Solve Ancient Questions: A Multi-Omics Study of the Stomach Content of the Oldest Human Ice Mummy, the 5300-Year-Old Iceman or Oetzi. In *Proteomic Profiling: Methods and Protocols*; Posch, A., Ed.; Springer US: New York, NY, 2021; pp 1–12 DOI: 10.1007/978-1-0716-1186-9_1.

(421) O'Connor, S.; Ali, E.; Al-Sabah, S.; Anwar, D.; Bergström, E.; Brown, K. A.; Buckberry, J.; Buckley, S.; Collins, M.; Denton, J.; et al. Exceptional Preservation of a Prehistoric Human Brain from Heslington, Yorkshire, UK. *J. Archaeol. Sci.* 2011, 38, 1641–1654.

(422) Morton-Hayward, A. L.; Thompson, T.; Thomas-Oates, J. E.; Buckley, S.; Petzold, A.; Ramsøe, A.; O'Connor, S.; Collins, M. J. A Conscious Rethink: Why Is Brain Tissue Commonly Preserved in the Archaeological Record? Commentary on: Petrone P; Pucci; P; Niola, M; et Al. Heat-Induced Brain Vitrification from the Vesuvius Eruption in C.E. 79. *Sci. Technol. Archaeol. Res.* 2020, 6, 87–95.

(423) Maixner, F.; Overath, T.; Linke, D.; Janko, M.; Guerriero, G.; van den Berg, B. H. J.; Stade, B.; Leidinger, P.; Backes, C.; Jaremek, M.; et al. Paleoproteomic Study of the Iceman's Brain Tissue. *Cell. Mol. Life Sci.* 2013, 70, 3709–3722.

(424) Petzold, A.; Lu, C.-H.; Groves, M.; Gobom, J.; Zetterberg, H.; Shaw, G.; O'Connor, S. Protein Aggregate Formation Permits Millennium-Old Brain Preservation. *J. R. Soc. Interface* 2020, 17, 20190775.

(425) Petrone, P.; Pucci, P.; Niola, M.; Baxter, P. J.; Fontanarosa, C.; Giordano, G.; Graziano, V.; Sirano, F.; Amoresano, A. Heat-Induced Brain Vitrification from the Vesuvius Eruption in C.e. 79. *N. Engl. J. Med.* 2020, 382, 383–384.

(426) Gallagher, D. E. Formation Processes of the Macrobotanical Record. In *Advances in Archaeological Method and Theory*; Academic Press, 2014; Vol. 10, pp 211–247. DOI: 10.1016/B978-0-12-003110-8.50007-4.

(427) Galland, M.; Job, D.; Rajjou, L. The Seed Proteome Web Portal. *Front. Plant Sci.* 2012, 3, 98.

(428) Mahalingam, R. Shotgun Proteomics of the Barley Seed Proteome. *BMC Genomics* 2017, 18, 1–11.

(429) Cappellini, E.; Gilbert, M. T. P.; Geuna, F.; Fiorentino, G.; Hall, A.; Thomas-Oates, J.; Ashton, P. D.; Ashford, D. A.; Arthur, P.; Campos, P. F.; et al. A Multidisciplinary Study of Archaeological Grape Seeds. *Naturwissenschaften* 2010, 97, 205–217.

(430) O'Donoghue, K.; Clapham, A.; Evershed, R. P.; Brown, T. A. Remarkable Preservation of Biomolecules in Ancient Radish Seeds. *Proc. Biol. Sci.* 1996, 263, 541–547.

(431) Mccobb, L. M. E.; Briggs, D. E. G.; Evershed, R. P.; Hall, A. R.; Hall, R. A. Preservation of Fossil Seeds From a 10th Century AD Cess Pit at Coppergate, York. *J. Archaeol. Sci.* 2001, 28, 929–940.

(432) Fiddyment, S.; Goodison, N. J.; Brenner, E.; Signorello, S.; Price, K.; Collins, M. J. Girding the Loins? Direct Evidence of the Use of a Medieval English Parchment Birthing Girdle from Biomolecular Analysis. *R. Soc. Open Sci.* 2021, 8, 202055.

(433) Maron, P.-A.; Ranjard, L.; Mougel, C.; Lemanceau, P. Metaproteomics: A New Approach for Studying Functional Microbial Ecology. *Microb. Ecol.* 2007, 53, 486–493.

(434) Consortium, The Human Microbiome Project. Structure, Function and Diversity of the Healthy Human Microbiome. *Nature* 2012, 486, 207–214.

(435) Eren, A. M.; Borisy, G. G.; Huse, S. M.; Mark Welch, J. L. Oligotyping Analysis of the Human Oral Microbiome. *Proc. Natl. Acad. Sci. U. S. A.* 2014, 111, E2875–E2884.

(436) Walters, K. E.; Martiny, J. B. H. Alpha-, Beta-, and Gamma-Diversity of Bacteria Varies across Habitats. *PLoS One* 2020, 15, No. e0233872.

(437) Warinner, C.; Herbig, A.; Mann, A.; Fellows Yates, J. A.; Weiß, C. L.; Burbano, H. A.; Orlando, L.; Krause, J. A Robust Framework for Microbial Archaeology. *Annu. Rev. Genomics Hum. Genet.* 2017, 18, 321–356.

(438) 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. London B Biol. Sci.* 2015, 370, 20130376.

(439) Mackie, M.; Hendy, J.; Lowe, A. D.; Sperduti, A.; Holst, M.; Collins, M. J.; Speller, C. F. Preservation of the Metaproteome: Variability of Protein Preservation in Ancient Dental Calculus. *Sci. Technol. Archaeol. Res.* 2017, 3, 74–86.

(440) Jersie-Christensen, R. R.; Lanigan, L. T.; Lyon, D.; Mackie, M.; Belstrøm, D.; Kelstrup, C. D.; Fotakis, A. K.; Willerslev, E.; Lynnerup, N.; Jensen, L. J.; et al. Quantitative Metaproteomics of Medieval Dental Calculus Reveals Individual Oral Health Status. *Nat. Commun.* 2018, 9, 1–12.

(441) Srinivasan, P. C. Neutrophils - the Sentinels of Periodontal Innate Immunity. *J. Clin. Cell. Immunol. S* 2013, 13, 1–6.

(442) Ryder, M. I. Comparison of Neutrophil Functions in Aggressive and Chronic Periodontitis. *Periodontol 2000* 2010, 53, 124–137.

(443) Charlton, S.; Ramsøe, A.; Collins, M.; Craig, O. E.; Fischer, R.; Alexander, M.; Speller, C. F. New Insights into Neolithic Milk Consumption through Proteomic Analysis of Dental Calculus. *Archaeol. Anthropol. Sci.* 2019, 11, 6183–6196.

(444) Geber, J.; Tromp, M.; Scott, A.; Bouwman, A.; Nanni, P.; Grossmann, J.; Hendy, J.; Warinner, C. Relief Food Subsistence Revealed by Microparticle and Proteomic Analyses of Dental Calculus from Victims of the Great Irish Famine. *Proc. Natl. Acad. Sci. U. S. A.* 2019, 116, 19380–19385.

(445) Jeong, C.; Wilkin, S.; Amgalantugs, T.; Bouwman, A. S.; Taylor, W. T. T.; Hagan, R. W.; Bromage, S.; Tsolmon, S.; Trachsel, C.; Grossmann, J.; et al. Bronze Age Population Dynamics and the Rise of Dairy Pastoralism on the Eastern Eurasian Steppe. *Proc. Natl. Acad. Sci. U. S. A.* 2018, 115, E11248–E11255.

(446) Runge, A. K. W.; Hendy, J.; Richter, K. K.; Masson-MacLean, E.; Britton, K.; Mackie, M.; McGrath, K.; Collins, M.; Cappellini, E.; Speller, C. Palaeoproteomic Analyses of Dog Palaeofaeces Reveal a Preserved Dietary and Host Digestive Proteome. *Proc. Biol. Sci.* 2021, 288, 20210020.

(447) Santiago-Rodriguez, T. M.; Narganes-Storde, Y. M.; Chanlatte, L.; Crespo-Torres, E.; Toranzos, G. A.; Jimenez-Flores, R.; Hamrick, A.; Cano, R. J. Microbial Communities in Pre-Columbian Coprolites. *PLoS One* 2013, 8, No. e65191.

(448) Gonçalves, M. L. C.; Araújo, A.; Duarte, R.; da Silva, J. P.; Reinhard, K.; Bouchet, F.; Ferreira, L. F. Detection of *Giardia Duodenalis* Antigen in Coprolites Using a Commercially Available Enzyme-Linked Immunosorbent Assay. *Trans. R. Soc. Trop. Med. Hyg.* 2002, 96, 640–643.

(449) Gonçalves, M. L. C.; Araújo, A.; Ferreira, L. F. Human Intestinal Parasites in the Past: New Findings and a Review. *Mem. Inst. Oswaldo Cruz* 2003, 98, 103−118.

(450) Goncalves, M. L. C.; da Silva, V. L.; de Andrade, C. M.; Reinhard, K.; da Rocha, G. C.; Le Bailly, M.; Bouchet, F.; Ferreira, L. F.; Araujo, A. Amoebiasis Distribution in the Past: First Steps Using an Immunoassay Technique. *Trans. R. Soc. Trop. Med. Hyg.* 2004, 98, 88−91.

(451) Mitchell, P. D.; Stern, E.; Tepper, Y. Dysentery in the Crusader Kingdom of Jerusalem: An ELISA Analysis of Two Medieval Latrines in the City of Acre (Israel). *J. Archaeol. Sci.* 2008, 35, 1849−1853.

(452) Warinner, C.; Speller, C.; Collins, M. J.; Lewis, C. M., Jr. Ancient Human Microbiomes. *J. Hum. Evol.* 2015, 79, 125−136.

(453) Shillito, L.-M.; Blong, J. C.; Green, E. J.; Van Asperen, E. The What, How and Why of Archaeological Coprolite Analysis. *Earth-Sci. Rev.* 2020, 207, 103196.

(454) Shoda, S.; Lucquin, A.; Ahn, J.-H.; Hwang, C.-J.; Craig, O. E. Pottery Use by Early Holocene Hunter-Gatherers of the Korean Peninsula Closely Linked with the Exploitation of Marine Resources. *Quat. Sci. Rev.* 2017, 170, 164−173.

(455) Craig, O. E.; Saul, H.; Lucquin, A.; Nishida, Y.; Taché, K.; Clarke, L.; Thompson, A.; Altoft, D. T.; Uchiyama, J.; Ajimoto, M.; et al. Earliest Evidence for the Use of Pottery. *Nature* 2013, 496, 351−354.

(456) Craig, O. E.; Forster, M.; Andersen, S. H.; Koch, E.; Crombé, P.; Milner, N. J.; Stern, B.; Bailey, G. N.; Heron, C. P. Molecular and Isotopic Demonstration of the Processing of Aquatic Products in Northern European Prehistoric Pottery. *Archaeometry* 2007, 49, 135−152.

(457) Heron, C.; Shoda, S.; Breu Barcons, A.; Czebreszuk, J.; Eley, Y.; Gorton, M.; Kirleis, W.; Kneisel, J.; Lucquin, A.; Müller, J.; et al. First Molecular and Isotopic Evidence of Millet Processing in Prehistoric Pottery Vessels. *Sci. Rep.* 2016, 6, 1−9.

(458) Evershed, R. P.; Payne, S.; Sherratt, A. G.; Copley, M. S.; Coolidge, J.; Urem-Kotsu, D.; Kotsakis, K.; Ozdoğan, M.; Ozdoğan, A. E.; Nieuwenhuyse, O.; et al. Earliest Date for Milk Use in the Near East and Southeastern Europe Linked to Cattle Herding. *Nature* 2008, 455, 528−531.

(459) Grillo, K. M.; Dunne, J.; Marshall, F.; Prendergast, M. E.; Casanova, E.; Gidna, A. O.; Janzen, A.; Karega-Munene Keute, J.; Mabulla, A. Z. P.; et al. Molecular and Isotopic Evidence for Milk, Meat, and Plants in Prehistoric Eastern African Herder Food Systems. *Proc. Natl. Acad. Sci. U. S. A.* 2020, 117, 9793−9799.

(460) Salque, M.; Bogucki, P. I.; Pyzel, J.; Sobkowiak-Tabaka, I.; Grygiel, R.; Szmyt, M.; Evershed, R. P. Earliest Evidence for Cheese Making in the Sixth Millennium BC in Northern Europe. *Nature* 2013, 493, 522−525.

(461) Romanus, K.; Baeten, J.; Poblome, J.; Accardo, S.; Degryse, P.; Jacobs, P.; De Vos, D.; Waelkens, M. Wine and Olive Oil Permeation in Pitched and Non-Pitched Ceramics: Relation with Results from Archaeological Amphorae from Sagalassos, Turkey. *J. Archaeol. Sci.* 2009, 36, 900−909.

(462) Rageot, M.; Mötsch, A.; Schorer, B.; Bardel, D.; Winkler, A.; Sacchetti, F.; Chaume, B.; Della Casa, P.; Buckley, S.; Cafisso, S.; et al. New Insights into Early Celtic Consumption Practices: Organic Residue Analyses of Local and Imported Pottery from Vix-Mont Lassois. *PLoS One* 2019, 14, No. e0218001.

(463) Drieu, L.; Rageot, M.; Wales, N.; Stern, B.; Lundy, J.; Zerrer, M.; Gaffney, I.; Bondetti, M.; Spiteri, C.; Thomas-Oates, J.; et al. Is It Possible to Identify Ancient Wine Production Using Biomolecular Approaches? *Sci. Technol.* 2020, 6, 16−29.

(464) McGovern, P.; Jalabadze, M.; Batiuk, S.; Callahan, M. P.; Smith, K. E.; Hall, G. R.; Kvavadze, E.; Maghradze, D.; Rusishvili, N.; Bouby, L.; et al. Early Neolithic Wine of Georgia in the South Caucasus. *P. Natl. Acad. Sci. U.S.A.* 2017, 114, E10309−E10318.

(465) Craig, O. E.; Taylor, G.; Mulville, J.; Collins, M. J.; Parker Pearson, M. The Identification of Prehistoric Dairying Activities in

the Western Isles of Scotland: An Integrated Biomolecular Approach. *J. Archaeol. Sci.* 2005, 32, 91−103.

(466) Craig, O. E.; Mulville, J.; Pearson, M. P.; Sokol, R.; Gelsthorpe, K.; Stacey, R.; Collins, M. Archaeology: Detecting Milk Proteins in Ancient Pots. *Nature* 2000, 408, 312−312.

(467) Solazzo, C.; Fitzhugh, W. W.; Rolando, C.; Tokarski, C. Identification of Protein Remains in Archaeological Potsherds by Proteomics. *Anal. Chem.* 2008, 80, 4590−4597.

(468) Craig, O. E.; Collins, M. J. An Improved Method for the Immunological Detection of Mineral Bound Protein Using Hydrofluoric Acid and Direct Capture. *J. Immunol. Methods* 2000, 236, 89−97.

(469) Amkreutz, L. W. S. *Persistent Traditions: A Long-Term Perspective on Communities in the Process of Neolithisation in the Lower Rhine Area (5500−2500 Cal BC)*; Sidestone Press, 2013.

(470) Nielsen, N. H.; Henriksen, P. S.; Mortensen, M. F.; Enevold, R.; Mortensen, M. N.; Scavenius, C.; Enghild, J. J. The Last Meal of Tollund Man: New Analyses of His Gut Content. *Antiquity* 2021, 95, 1195−1212.

(471) Bos, K. I.; Kühnert, D.; Herbig, A.; Esquivel-Gomez, L. R.; Andrades Valtueña, A.; Barquera, R.; Giffin, K.; Kumar Lankapalli, A.; Nelson, E. A.; Sabin, S.; et al. Paleomicrobiology: Diagnosis and Evolution of Ancient Pathogens. *Annu. Rev. Microbiol.* 2019, 73, 639−666.

(472) Spyrou, M. A.; Bos, K. I.; Herbig, A.; Krause, J. Ancient Pathogen Genomics as an Emerging Tool for Infectious Disease Research. *Nat. Rev. Genet.* 2019, 20, 323−340.

(473) Boros-Major, A.; Bona, A.; Lovasz, G.; Molnar, E.; Marcsik, A.; Palfi, G.; Mark, L. New Perspectives in Biomolecular Paleopathology of Ancient Tuberculosis: A Proteomic Approach. *J. Archaeol. Sci.* 2011, 38, 197−201.

(474) Hajdu, T.; Fóthi, E.; Kovári, I.; Merczi, M.; Molnár, A.; Maász, G.; Avar, P.; Marcsik, A.; Márk, L. Bone Tuberculosis in Roman Period Pannonia (Western Hungary). *Mem. Inst. Oswaldo Cruz* 2012, 107, 1048−1053.

(475) Hendy, J.; Collins, M.; Teoh, K. Y.; Ashford, D. A.; Thomas-Oates, J.; Donoghue, H. D.; Pap, I.; Minnikin, D. E.; Spigelman, M.; Buckley, M. The Challenge of Identifying Tuberculosis Proteins in Archaeological Tissues. *J. Archaeol. Sci.* 2016, 66, 146−153.

(476) Schmidt-Schultz, T. H.; Schultz, M. AG 85, a Major Secretion Protein of *Mycobacterium Tuberculosis*, Can Be Identified in Ancient Bone. *Tuberculosis* 2015, 95, S87−S92.

(477) Fotakis, A. K.; Denham, S. D.; Mackie, M.; Orbegozo, M. I.; Mylopotamitaki, D.; Gopalakrishnan, S.; Sicheritz-Pontén, T.; Olsen, J. V.; Cappellini, E.; Zhang, G.; et al. Multi-Omic Detection of *Mycobacterium Leprae* in Archaeological Human Dental Calculus. *Philos. Trans. R. Soc. London B Biol. Sci.* 2020, 375, 20190584.

(478) Bos, K. I.; Schuenemann, V. J.; Golding, B.; Burbano, H. A.; Waglechner, N.; Coombes, B. K.; McPhee, J. B.; DeWitte, S. N.; Meyer, M.; Schmedes, S.; et al. A Draft Genome of *Yersinia Pestis* from Victims of the Black Death. *Nature* 2011, 478, 506−510.

(479) Dallongeville, S.; Garnier, N.; Rolando, C.; Tokarski, C. Proteins in Art, Archaeology, and Paleontology: From Detection to Identification. *Chem. Rev.* 2016, 116, 2−79.

(480) Vilanova, C.; Porcar, M. Art-omics: Multi-omics Meet Archaeology and Art Conservation. *Microb. Biotechnol.* 2020, 13, 435−441.

(481) Kuckova, S.; Cejnar, P.; Santrucek, J.; Hynek, R. Characterization of Proteins in Cultural Heritage Using MALDI-TOF and LC−MS/MS Mass Spectrometric Techniques. *Phys. Sci. Rev.* 2019, 4, 11−15.

(482) Giuffrida, M. G.; Mazzoli, R.; Pessione, E. Back to the Past: Deciphering Cultural Heritage Secrets by Protein Identification. *Appl. Microbiol. Biotechnol.* 2018, 102, 5445−5455.

(483) Vinciguerra, R.; De Chiaro, A.; Pucci, P.; Marino, G.; Birolo, L. Proteomic Strategies for Cultural Heritage: From Bones to Paintings. *Microchem. J.* 2016, 126, 341−348.

(484) Tripković, T.; Charvy, C.; Alves, S.; Lolić, A. Đ.; Baošić, R. M.; Nikolić-Mandić, S. D.; Tabet, J. C. Identification of Protein

Binders in Artworks by MALDI-TOF/TOF Tandem Mass Spectrometry. *Talanta* **2013**, *113*, 49−61.

(485) Hynek, R.; Kuckova, S.; Hradilova, J.; Kodicek, M. 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 Commun. Mass Spectrom.* **2004**, *18*, 1896−1900.

(486) Kuckova, S.; Hynek, R.; Kodicek, M. Identification of Proteinaceous Binders Used in Artworks by MALDI-TOF Mass Spectrometry. *Anal. Bioanal. Chem.* **2007**, *388*, 201−206.

(487) Leo, G.; Cartechini, L.; Pucci, P.; Sgamellotti, A.; Marino, G.; Birolo, L. Proteomic Strategies for the Identification of Proteinaceous Binders in Paintings. *Anal. Bioanal. Chem.* **2009**, *395*, 2269−2280.

(488) Pozzi, F.; Arslanoglu, J.; Galluzzi, F.; Tokarski, C.; Snyder, R. Mixing, Dipping, and Fixing: The Experimental Drawing Techniques of Thomas Gainsborough. *Herit. Sci.* **2020**, *8*, 1−14.

(489) Levy, I. K.; Neme Tauil, R.; Rosso, A.; Valacco, M. P.; Moreno, S.; Guzmán, F.; Siracusano, G.; Maier, M. S. Finding of Muscle Proteins in Art Samples from Mid-18th Century Murals by LC−MSMS. *J. Cult. Herit.* **2021**, *48*, 227−235.

(490) Calvano, C. D.; Rigante, E.; Picca, R. A.; Cataldi, T. R. I.; Sabbatini, L. An Easily Transferable Protocol for in-Situ Quasi-Non-Invasive Analysis of Protein Binders in Works of Art. *Talanta* **2020**, *215*, 120882.

(491) Calvano, C. D.; van der Werf, I. D.; Palmisano, F.; Sabbatini, L. Identification of Lipid- and Protein-Based Binders in Paintings by Direct On-Plate Wet Chemistry and Matrix-Assisted Laser Desorption Ionization Mass Spectrometry. *Anal. Bioanal. Chem.* **2015**, *407*, 1015−1022.

(492) Kuckova, S.; Sandu, I. C. A.; Crhova, M.; Hynek, R.; Fogas, I.; Schafer, S. Protein Identification and Localization Using Mass Spectrometry and Staining Tests in Cross-Sections of Polychrome Samples. *J. Cult. Herit.* **2013**, *14*, 31−37.

(493) van der Werf, I. D.; Calvano, C. D.; Palmisano, F.; Sabbatini, L. A Simple Protocol for Matrix Assisted Laser Desorption Ionization - Time of Flight-Mass Spectrometry (MALDI-TOF-MS) Analysis of Lipids and Proteins in Single Microsamples of Paintings. *Anal. Chim. Acta* **2012**, *718*, 1−10.

(494) Fremout, W.; Dhaenens, M.; Saverwyns, S.; Sanyova, J.; Vandenabeele, P.; Deforce, D.; Moens, L. Tryptic Peptide Analysis of Protein Binders in Works of Art by Liquid Chromatography-Tandem Mass Spectrometry. *Anal. Chim. Acta* **2010**, *658*, 156−162.

(495) Chambery, A.; Di Maro, A.; Sanges, C.; Severino, V.; Tarantino, M.; Lamberti, A.; Parente, A.; Arcari, P. Improved Procedure for Protein Binder Analysis in Mural Painting by LC-ESI/Q-Q-TOF Mass Spectrometry: Detection of Different Milk Species by Casein Proteotypic Peptides. *Anal. Bioanal. Chem.* **2009**, *395*, 2281−2291.

(496) Tripković, T.; Charvy, C.; Alves, S.; Lolić, A. Đ.; Baošić, R. M.; Nikolić-Mandić, S. D.; Tabet, J. C. Electrospray Ionization Linear Trap Quadrupole Orbitrap in Analysis of Old Tempera Paintings: Application to Nineteenth-Century Orthodox Icons. *Eur. J. Mass Spectrom.* **2015**, *21*, 679−692.

(497) Levy, I. K.; Neme Tauil, R.; Valacco, M. P.; Moreno, S.; Siracusano, G.; Maier, M. S. Investigation of Proteins in Samples of a Mid-18th Century Colonial Mural Painting by MALDI-TOF/MS and LC-ESI/MS (Orbitrap). *Microchem. J.* **2018**, *143*, 457−466.

(498) 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* **2015**, *7*, 143−149.

(499) Kuckova, S.; Crhova, M.; Vankova, L.; Hnizda, A.; Hynek, R.; Kodicek, M. Towards Proteomic Analysis of Milk Proteins in Historical Building Materials. *Int. J. Mass Spectrom.* **2009**, *284*, 42−46.

(500) Krizova, I.; Schultz, J.; Nemec, I.; Cabala, R.; Hynek, R.; Kuckova, S. Comparison of Analytical Tools Appropriate for Identification of Proteinaceous Additives in Historical Mortars. *Anal. Bioanal. Chem.* **2018**, *410*, 189−200.

(501) Righetti, P. G.; Zilberstein, G.; Zilberstein, S. New Baits for Fishing in Cultural Heritage's *Mare Magnum*. *J. Proteomics* **2021**, *235*, 104113.

(502) Ntasi, G.; Kirby, D. P.; Stanzione, I.; Carpentieri, A.; Somma, P.; Cicatiello, P.; Marino, G.; Giardina, P.; Birolo, L. A Versatile and User-Friendly Approach for the Analysis of Proteins in Ancient and Historical Objects. *J. Proteomics* **2021**, *231*, 104039.

(503) Kirby, D. P.; Manick, A.; Newman, R. Minimally Invasive Sampling of Surface Coatings for Protein Identification by Peptide Mass Fingerprinting: A Case Study with Photographs. *J. Am. Inst. Conserv.* **2020**, *59*, 235−245.

(504) Bleicher, N.; Kelstrup, C.; Olsen, J. V.; Cappellini, E. Molecular Evidence of Use of Hide Glue in 4th Millennium BC Europe. *J. Archaeol. Sci.* **2015**, *63*, 65−71.

(505) Rao, H.; Yang, Y.; Abuduresule, I.; Li, W.; Hu, X.; Wang, C. Proteomic Identification of Adhesive on a Bone Sculpture-Inlaid Wooden Artifact from the Xiaohe Cemetery, Xinjiang, China. *J. Archaeol. Sci.* **2015**, *53*, 148−155.

(506) Mai, H.; Yang, Y.; Abuduresule, I.; Li, W.; Hu, X.; Wang, C. Characterization of Cosmetic Sticks at Xiaohe Cemetery in Early Bronze Age Xinjiang, China. *Sci. Rep.* **2016**, *6*, 1−9.

(507) Popowich, A. K.; Cleland, T. P.; Solazzo, C. Characterization of Membrane Metal Threads by Proteomics and Analysis of a 14th C. Thread from an Italian Textile. *J. Cult. Herit.* **2018**, *33*, 10−17.

(508) Solazzo, C.; Courel, B.; Connan, J.; van Dongen, B. E.; Barden, H.; Penkman, K.; Taylor, S.; Demarchi, B.; Adam, P.; Schaeffer, P.; et al. Identification of the Earliest Collagen-and-Plant-Based Coatings from Neolithic Artefacts (Nahal Hemar Cave, Israel). *Sci. Rep.* **2016**, *6*, 1−11.

(509) Galluzzi, F. *New Developments in High-Resolution Mass Spectrometry for Proteomic Analysis Applied to Cultural Heritage: Study of Intact Proteins, Their Cross-Linking and Interaction in Artworks and Museum Objects, Bordeaux*, **2021**.

(510) Zilberstein, G.; Zilberstein, S.; Maor, U.; Righetti, P. G. Surface Analysis of Ancient Parchments via the EVA Film: The Aleppo Codex. *Anal. Biochem.* **2020**, *604*, 113824.

(511) Hao, P.; Adav, S. S.; Gallart-Palau, X.; Sze, S. K. Recent Advances in Mass Spectrometric Analysis of Protein Deamidation. *Mass Spectrom. Rev.* **2017**, *36*, 677−692.

(512) Palmer, K. S.; Makarewicz, C. A.; Tishkin, A. A.; Tur, S. S.; Chunag, A.; Diimajav, E.; Jamsranjav, B.; Buckley, M. Comparing the Use of Magnetic Beads with Ultrafiltration for Ancient Dental Calculus Proteomics. *J. Proteome Res.* **2021**, *20*, 1689−1704.

(513) Kelstrup, C. D.; Bekker-Jensen, D. B.; Arrey, T. N.; Hogrebe, A.; Harder, A.; Olsen, J. V. Performance Evaluation of the Q Exactive HF-X for Shotgun Proteomics. *J. Proteome Res.* **2018**, *17*, 727−738.

(514) Kreimer, S.; Belov, M. E.; Danielson, W. F.; Levitsky, L. I.; Gorshkov, M. V.; Karger, B. L.; Ivanov, A. R. Advanced Precursor Ion Selection Algorithms for Increased Depth of Bottom-up Proteomic Profiling. *J. Proteome Res.* **2016**, *15*, 3563−3573.

(515) Gao, Z.; Chang, C.; Yang, J.; Zhu, Y.; Fu, Y. AP3: An Advanced Proteotypic Peptide Predictor for Targeted Proteomics by Incorporating Peptide Digestibility. *Anal. Chem.* **2019**, *91*, 8705−8711.

(516) Serrano, G.; Guruceaga, E.; Segura, V. DeepMSPeptide: Peptide Detectability Prediction Using Deep Learning. *Bioinformatics* **2020**, *36*, 1279−1280.

(517) Mallick, P.; Schirle, M.; Chen, S. S.; Flory, M. R.; Lee, H.; Martin, D.; Ranish, J.; Raught, B.; Schmitt, R.; Werner, T.; et al. Computational Prediction of Proteotypic Peptides for Quantitative Proteomics. *Nat. Biotechnol.* **2007**, *25*, 125−131.

(518) Schubert, O. T.; Röst, H. L.; Collins, B. C.; Rosenberger, G.; Aebersold, R. Quantitative Proteomics: Challenges and Opportunities in Basic and Applied Research. *Nat. Protoc.* **2017**, *12*, 1289−1294.

(519) Mergner, J.; Frejno, M.; List, M.; Papacek, M.; Chen, X.; Chaudhary, A.; Samaras, P.; Richter, S.; Shikata, H.; Messerer, M.; et al. Mass Spectrometry-Based Draft of the *Arabidopsis* Proteome. *Nature* **2020**, *579*, 409−414.

(520) Schaller, J.; Gerber, S.; Kaempfer, U.; Lejon, S.; Trachsel, C. *Human Blood Plasma Proteins: Structure and Function*; John Wiley & Sons, 2008.

(521) Hortin, G. L.; Sviridov, D. The Dynamic Range Problem in the Analysis of the Plasma Proteome. *J. Proteomics* **2010**, *73*, 629−636.

(522) Sudmant, P. H.; Kitzman, J. O.; Antonacci, F.; Alkan, C.; Malig, M.; Tsalenko, A.; Sampas, N.; Bruhn, L.; Shendure, J. 1000 Genomes Project; Eichler, E. E.; Diversity of Human Copy Number Variation and Multicopy Genes. *Science* **2010**, *330*, 641−646.

(523) Egertson, J. D.; Kuehn, A.; Merrihew, G. E.; Bateman, N. W.; MacLean, B. X.; Ting, Y. S.; Canterbury, J. D.; Marsh, D. M.; Kellmann, M.; Zabrouskov, V.; et al. Multiplexed MS/MS for Improved Data-Independent Acquisition. *Nat. Methods* **2013**, *10*, 744−746.

(524) Colangelo, C. M.; Chung, L.; Bruce, C.; Cheung, K.-H. Review of Software Tools for Design and Analysis of Large Scale MRM Proteomic Datasets. *Methods* **2013**, *61*, 287−298.

(525) Meier, F.; Geyer, P. E.; Virreira Winter, S.; Cox, J.; Mann, M. BoxCar Acquisition Method Enables Single-Shot Proteomics at a Depth of 10,000 Proteins in 100 minutes. *Nat. Methods* **2018**, *15*, 440−448.

(526) Zhang, F.; Ge, W.; Ruan, G.; Cai, X.; Guo, T. Data-Independent Acquisition Mass Spectrometry-Based Proteomics and Software Tools: A Glimpse in 2020. *Proteomics* **2020**, *20*, No. e1900276.

(527) Denisov, E.; Damoc, E.; Makarov, A. Exploring Frontiers of Orbitrap Performance for Long Transients. *Int. J. Mass Spectrom.* **2021**, *466*, 116607.

(528) Révész, Á.; Milley, M. G.; Nagy, K.; Szabó, D.; Kalló, G.; Csősz, É.; Vékey, K.; Drahos, L. Tailoring to Search Engines: Bottom-up Proteomics with Collision Energies Optimized for Identification Confidence. *J. Proteome Res.* **2021**, *20*, 474−484.

(529) Guan, S.; Taylor, P. P.; Han, Z.; Moran, M. F.; Ma, B. Data Dependent-Independent Acquisition (DDIA) Proteomics. *J. Proteome Res.* **2020**, *19*, 3230−3237.

(530) Gao, Y.; Ping, L.; Duong, D.; Zhang, C.; Dammer, E. B.; Li, Y.; Chen, P.; Chang, L.; Gao, H.; Wu, J.; Xu, P. Mass Spectrometry-Based Near-Complete Draft of the *Saccharomyces Cerevisiae* Proteome. *J. Proteome Res.* **2021**, *20*, 1328−1340.

(531) Narayan, V.; Ly, T.; Pourkarimi, E.; Murillo, A. B.; Gartner, A.; Lamond, A. I.; Kenyon, C. Deep Proteome Analysis Identifies Age-Related Processes in *C. Elegans*. *Cell Syst.* **2016**, *3*, 144−159.

(532) Lewin, H. A.; Robinson, G. E.; Kress, W. J.; Baker, W. J.; Coddington, J.; Crandall, K. A.; Durbin, R.; Edwards, S. V.; Forest, F.; Gilbert, M. T. P.; et al. Earth BioGenome Project: Sequencing Life for the Future of Life. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, 4325−4333.

(533) Rhie, A.; McCarthy, S. A.; Fedrigo, O.; Damas, J.; Formenti, G.; Koren, S.; Uliano-Silva, M.; Chow, W.; Fungtammasan, A.; Kim, J.; et al. Towards Complete and Error-Free Genome Assemblies of All Vertebrate Species. *Nature* **2021**, *592*, 737−746.

(534) Genome 10K Community of Scientists. Genome 10K: A Proposal to Obtain Whole-Genome Sequence for 10 000 Vertebrate Species. *J. Hered.* **2009**, *100*, 659−674.

(535) Darwin Tree of Life Project Consortium. Sequence Locally, Think Globally: The Darwin Tree of Life Project. *Proc. Natl. Acad. Sci. U. S. A.* **2022**, *119*, No. e2115642118.

(536) Richter, K. K.; Wilson, J.; Jones, A. K. G.; Buckley, M.; van Doorn, N.; Collins, M. J. Fish 'n Chips: ZooMS Peptide Mass Fingerprinting in a 96 Well Plate Format to Identify Fish Bone Fragments. *J. Archaeol. Sci.* **2011**, *38*, 1502−1510.

(537) Deutsch, E. W.; Bandeira, N.; Sharma, V.; Perez-Riverol, Y.; Carver, J. J.; Kundu, D. J.; García-Seisdedos, D.; Jarnuczak, A. F.; Hewapathirana, S.; Pullman, B. S.; et al. The ProteomeXchange Consortium in 2020: Enabling "Big Data" Approaches in Proteomics. *Nucleic Acids Res.* **2019**, *48*, D1145−D1152.

(538) Perez-Riverol, Y.; Bai, J.; Bandla, C.; García-Seisdedos, D.; Hewapathirana, S.; Kamatchinathan, S.; Kundu, D. J.; Prakash, A.;

Frericks-Zipper, A.; Eisenacher, M.; et al. The PRIDE Database Resources in 2022: A Hub for Mass Spectrometry-Based Proteomics Evidences. *Nucleic Acids Res.* **2022**, *50*, D543−D552.

(539) Tsutaya, T.; Mackie, M.; Koenig, C.; Sato, T.; Weber, A. W.; Kato, H.; Olsen, J. V.; Cappellini, E. Palaeoproteomic Identification of Breast Milk Protein Residues from the Archaeological Skeletal Remains of a Neonatal Dog. *Sci. Rep.* **2019**, *9*, 12841.

(540) Medha, K.; Rajna, S.; Devi, L. J.; Samant, L.; Jose, S. A Comprehensive Review on Moth Repellent Finishing of Woolen Textiles. *J. Cult. Herit.* **2021**, *49*, 260−271.

(541) Elder, A.; Madsen, S.; Brown, G.; Herbel, C.; Collins, C.; Whelan, S.; Wenz, C.; Alderson, S.; Kronthal, L. Adhesives and Consolidants in Geological and Paleontological Conservation. Part One: Introduction, Guide, Health and Safety, Definitions. *SPNHC Leaflets* **1997**, *1*, 1−4.

(542) Johnson, J. S. Consolidation of Archaeological Bone: A Conservation Perspective. *J. Field Archaeol.* **1994**, *21*, 221−233.

(543) Schellmann, N. C. Animal Glues: A Review of Their Key Properties Relevant to Conservation. *Stud. Conserv.* **2007**, *52*, 55−66.

(544) Zhou, Y.; Zheng, H.-L.; Hu, Z.-W.; Zhao, F. Research on the Use of Fibroin Protein for Consolidation of Ancient Silk Textiles. *Sciences of Conservation and Archaeology* **2010**, *22*, 44−48.

(545) Liu, X.; Zhang, Q.; Peng, B.; Ren, Y.; Cheng, B.; Ding, C.; Su, X.; He, J.; Lin, S. Flame Retardant Cellulosic Fabrics via Layer-by-Layer Self-Assembly Double Coating with Egg White Protein and Phytic Acid. *J. Clean. Prod.* **2020**, *243*, 118641.

(546) Brock, F.; Dee, M.; Hughes, A.; Snoeck, C.; Staff, R.; Bronk Ramsey, C. Testing the Effectiveness of Protocols for Removal of Common Conservation Treatments for Radiocarbon Dating. *Radiocarbon* **2018**, *60*, 35−50.

(547) Eklund, J. A. The Possible Effects of Preparation and Conservation Treatments on DNA Research, Using Treatments for Basketry as a Case Study. *J. Am. Inst. Conserv.* **2012**, *35*, 92−102.

(548) Stephan, E. Oxygen Isotope Analysis of Animal Bone Phosphate: Method Refinement, Influence of Consolidants, and Reconstruction of Palaeotemperatures for Holocene Sites. *J. Archaeol. Sci.* **2000**, *27*, 523−535.

(549) Orlando, L.; Allaby, R.; Skoglund, P.; Sarkissian, C. D.; Stockhammer, P. W.; Ávila-Arcos, M. C.; Fu, Q.; Krause, J.; Willerslev, E.; Stone, A. C.; et al. Ancient DNA Analysis. *Nat. Rev. Dis. Primers* **2021**, *1*, 1−26.

(550) Hofreiter, M.; Serre, D.; Poinar, H. N.; Kuch, M.; Pääbo, S. Ancient DNA. *Nat. Rev. Genet.* **2001**, *2*, 353−359.

(551) Briggs, A. W.; Stenzel, U.; Johnson, P. L. F.; Green, R. E.; Kelso, J.; Prüfer, K.; Meyer, M.; Krause, J.; Ronan, M. T.; Lachmann, M.; et al. Patterns of Damage in Genomic DNA Sequences from a Neandertal. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 14616−14621.

(552) Jónsson, H.; Ginolhac, A.; Schubert, M.; Johnson, P. L. F.; Orlando, L. mapDamage2.0: Fast Approximate Bayesian Estimates of Ancient DNA Damage Parameters. *Bioinformatics* **2013**, *29*, 1682−1684.

(553) Skoglund, P.; Northoff, B. H.; Shunkov, M. V.; Derevianko, A. P.; Pääbo, S.; Krause, J.; Jakobsson, M. Separating Endogenous Ancient DNA from Modern Day Contamination in a Siberian Neandertal. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 2229−2234.

(554) Ramsøe, A.; van Heekeren, V.; Ponce, P.; Fischer, R.; Barnes, I.; Speller, C.; Collins, M. J. DeamiDATE 1.0: Site-Specific Deamidation as a Tool to Assess Authenticity of Members of Ancient Proteomes. *J. Archaeol. Sci.* **2020**, *115*, 105080.

(555) Schroeter, E. R.; Cleland, T. P. Glutamine Deamidation: An Indicator of Antiquity, or Preservational Quality? *Rapid Commun. Mass Spectrom.* **2016**, *30*, 251−255.

(556) 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.* **2012**, *26*, 2319−2327.

(557) Wilson, J.; van Doorn, N. L.; Collins, M. J. Assessing the Extent of Bone Degradation Using Glutamine Deamidation in Collagen. *Anal. Chem.* **2012**, *84*, 9041−9048.

(558) Welker, F.; Soressi, M. A.; Roussel, M.; van Riemsdijk, I.; Hublin, J.-J.; Collins, M. J. Variations in Glutamine Deamidation for a Châtelperronian Bone Assemblage as Measured by Peptide Mass Fingerprinting of Collagen. *Sci. Technol. Archaeol.* **2017**, *3*, 15−27.

(559) Ying, Y.; Li, H. Recent Progress in the Analysis of Protein Deamidation Using Mass Spectrometry. *Methods* **2022**, *200*, 42−57.

(560) Brown, S.; Kozlikin, M.; Shunkov, M.; Derevianko, A.; Higham, T.; Douka, K.; Richter, K. K. Examining Collagen Preservation through Glutamine Deamidation at Denisova Cave. *J. Archaeol. Sci.* **2021**, *133*, 105454.

(561) Simpson, J. P.; Penkman, K. E. H.; Demarchi, B.; Koon, H.; Collins, M. J.; Thomas-Oates, J.; Shapiro, B.; Stark, M.; Wilson, J. The Effects of Demineralisation and Sampling Point Variability on the Measurement of Glutamine Deamidation in Type I Collagen Extracted from Bone. *J. Archaeol. Sci.* **2016**, *69*, 29−38.

(562) Ramsøe, A.; Crispin, M.; Mackie, M.; McGrath, K.; Fischer, R.; Demarchi, B.; Collins, M. J.; Hendy, J.; Speller, C. Assessing the Degradation of Ancient Milk Proteins through Site-Specific Deamidation Patterns. *Sci. Rep.* **2021**, *11*, 7795.

(563) Pal Chowdhury, M.; Wogelius, R.; Manning, P. L.; Metz, L.; Slimak, L.; Buckley, M. Collagen Deamidation in Archaeological Bone as an Assessment for Relative Decay Rates. *Archaeometry* **2019**, *61*, 1382−1398.

(564) Schmidt-Schultz, T. H.; Schultz, M. Bone Protects Proteins over Thousands of Years: Extraction, Analysis, and Interpretation of Extracellular Matrix Proteins in Archaeological Skeletal Remains. *Am. J. Phys. Anthropol.* **2004**, *123*, 30−39.

(565) 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.* **2017**, *78*, 29−39.

(566) Bleasdale, M.; Boivin, N.; Richter, K. K. Oral Signature Screening Database for Palaeoproteomic Analyses of Dental Calculus. *zenodo.net* **2020**, DOI: 10.5281/zenodo.3698271.

(567) Perdigão, N.; Heinrich, J.; Stolte, C.; Sabir, K. S.; Buckley, M. J.; Tabor, B.; Signal, B.; Gloss, B. S.; Hammang, C. J.; Rost, B.; et al. Unexpected Features of the Dark Proteome. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 15898−15903.

(568) Meyer, J. G.; Niemi, N. M.; Pagliarini, D. J.; Coon, J. J. Quantitative Shotgun Proteome Analysis by Direct Infusion. *Nat. Methods* **2020**, *17*, 1222−1228.

(569) Pinheiro, F.; Santos, J.; Ventura, S. AlphaFold and the Amyloid Landscape. *J. Mol. Biol.* **2021**, *433*, 167059.

(570) DiDonato, N.; Chen, H.; Waggoner, D.; Hatcher, P. G. Potential Origin and Formation for Molecular Components of Humic Acids in Soils. *Geochim. Cosmochim. Acta* **2016**, *178*, 210−222.

(571) Ikeya, K.; Sleighter, R. L.; Hatcher, P. G.; Watanabe, A. Characterization of the Chemical Composition of Soil Humic Acids Using Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Geochim. Cosmochim. Acta* **2015**, *153*, 169−182.

(572) Sutton, P. A.; Lewis, C. A.; Rowland, S. J. Isolation of Individual Hydrocarbons from the Unresolved Complex Hydrocarbon Mixture of a Biodegraded Crude Oil Using Preparative Capillary Gas Chromatography. *Org. Geochem.* **2005**, *36*, 963−970.

(573) Reznikov, N.; Bilton, M.; Lari, L.; Stevens, M. M.; Kröger, R. Fractal-like Hierarchical Organization of Bone Begins at the Nanoscale. *Sci. Adv.* **2018**, DOI: 10.1126/sciadv.abd6495.

(574) Lundberg, E.; Borner, G. H. H. Spatial Proteomics: A Powerful Discovery Tool for Cell Biology. *Nat. Rev. Mol. Cell Biol.* **2019**, *20*, 285−302.

(575) Paul, I.; White, C.; Turcinovic, I.; Emili, A. Imaging the Future: The Emerging Era of Single-Cell Spatial Proteomics. *FEBS J.* **2021**, *288*, 6990−7001.

(576) Hale, O. J.; Cooper, H. J. Native Mass Spectrometry Imaging and In Situ Top-down Identification of Intact Proteins Directly from Tissue. *J. Am. Soc. Mass Spectrom.* **2020**, *31*, 2531−2537.

(577) Schermelleh, L.; Ferrand, A.; Huser, T.; Eggeling, C.; Sauer, M.; Biehlmaier, O.; Drummen, G. P. C. Super-Resolution Microscopy Demystified. *Nat. Cell Biol.* **2019**, *21*, 72−84.

(578) Callahan, N.; Tullman, J.; Kelman, Z.; Marino, J. Strategies for Development of a Next-Generation Protein Sequencing Platform. *Trends Biochem. Sci.* **2020**, *45*, 76−89.

(579) Howorka, S.; Siwy, Z. S. Reading Amino Acids in a Nanopore. *Nat. Biotechnol.* **2020**, *38*, 159−160.

(580) Murphy, K. R.; Timko, S. A.; Gonsior, M.; Powers, L. C.; Wünsch, U. J.; Stedmon, C. A. Photochemistry Illuminates Ubiquitous Organic Matter Fluorescence Spectra. *Environ. Sci. Technol.* **2018**, *52*, 11243−11250.

(581) Abelson, P. H. PALEOBIOCHEMISTRY. *Sci. Am.* **1956**, *195*, 83−96.

(582) Abelson, P. H. Some Aspects of Paleo-Biochemistry. *Ann. N.Y. Acad. Sci.* **1957**, *69*, 276−285.

(583) Abelson, P. H. Amino Acids in Fossils. *Science* **1954**, *119*, 576−576.

(584) Hare, P. E. The Amino Acid Composition of the Organic Matrix of Some Recent and Fossil Shells of Some West Coast Species of Mytilus. PhD Dissertation, California Institute of Technology, 1962. DOI: 10.7907/EXQX-4D49

(585) Hare, P. E. Effect of Hydrolysis on the Racemization Rate of Amino Acids. *Carnegie Institute of Washington Yearbook* **1971**, *70*, 256−258.

(586) Wehmiller, J.; Hare, P. E. Racemization of Amino Acids in Marine Sediments. *Science* **1971**, *173*, 907−911.

(587) Hare, P. E.; Mitterer, R. M. Non-Protein Amino Acids in Fossil Shells. *Carnegie Institution of Washington Yearbook* **1967**, *65*, 362−372.

(588) Hare, P. E.; Mitterer, R. M. Laboratory Simulation of Amino Acid Diagenesis in Fossils. *Carnegie Institute of Washington Yearbook* **1969**, *67*, 205−208.

(589) Hare, P. E.; Turnbull, H. F.; Taylor, R. E. Amino Acid Dating of Pleistocene Fossil Materials: Olduvai Gorge, Tanzania. In *Views of the Past*; Freeman, L., Ed.; De Gruyter Mouton, 1978; pp 7−12.

(590) Hare, P. E. Amino Acid Dating of Bone—the Influence of Water. *Carnegie Institute of Washington Yearbook* **1974**, *73*, 576−581.

(591) Bada, J. L. Amino Acid Racemization Dating of Fossil Bones. *Annu. Rev. Earth Planet. Sci.* **1985**, *13*, 241−268.

(592) Hoering, T. C.; Ford, H. T. The Isotope Effect in the Fixation of Nitrogen by Azotobacter. *J. Am. Chem. Soc.* **1960**, *82*, 376−378.

(593) Abelson, P. H.; Hoering, T. C. Carbon Isotope Fractionation in Formation of Amino Acids by Photosynthetic Organisms. *Proc. Natl. Acad. Sci. U. S. A.* **1961**, *47*, 623−632.

(594) Hoering, T. C.; Hare, P. E. Comparison of Natural Humic Acids with Amino Acid - Glucose Reaction Products. *AAPG Bull.* **1973**, *57*, 784−784.

(595) Jope, M. The Protein of Brachiopod Shell—I. Amino Acid Composition and Implied Protein Taxonomy. *Comp. Biochem. Physiol.* **1967**, *20*, 593−600.

(596) Muyzer, G.; Sandberg, P.; Knapen, M. H. J.; Vermeer, C.; Collins, M.; Westbroek, P. Preservation of the Bone Protein Osteocalcin in Dinosaurs. *Geology* **1992**, *20*, 871−874.

(597) Muyzer, G.; Westbroek, P.; Wehmiller, J. F. Phylogenetic Implications and Diagenetic Stability of Macromolecules from Pleistocene and Recent Shells of *Mercenaria Mercenaria* (Mollusca, Bivalvia). *Hist. Biol.* **1988**, *1*, 135−144.

(598) Westbroek, P. *Life as a Geological Force*; Norton, 1991.

(599) Ostrom, P. H.; Macko, S. A.; Engel, M. H.; Silfer, J. A.; Russell, D. Geochemical Characterization of High Molecular Weight Material Isolated from Late Cretaceous Fossils. *Org. Geochem.* **1990**, *16*, 1139−1144.

(600) Ostrom, P. H.; Macko, S. A.; Engel, M. H.; Russell, D. A. Assessment of Trophic Structure of Cretaceous Communities Based on Stable Nitrogen Isotope Analyses. *Geology* **1993**, *21*, 491−494.

(601) Robbins, L. L.; Ostrom, P. H. Molecular Isotopic and Biochemical Evidence of the Origin and Diagenesis of Shell Organic Material. *Geology* **1995**, *23*, 345−348.

(602) Ostrom, P. H.; Zonneveld, J.-P.; Robbins, L. L. Organic Geochemistry of Hard Parts: Assessment of Isotopic Variability and

Indigeneity. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **1994**, *107*, 201−212.

(603) Uskoković, V.; Khan, F.; Liu, H.; Witkowska, H. E.; Zhu, L.; Li, W.; Habelitz, S. Hydrolysis of Amelogenin by Matrix Metalloprotease-20 Accelerates Mineralization in Vitro. *Arch. Oral Biol.* **2011**, *56*, 1548−1559.

(604) *FishBase 2000: Concepts, Design and Data Sources.*; ICLARM: Los Baños, Laguna, Philippines, 2000.

(605) *IOC World Bird List (v11.2)*, 2021. DOI: 10.14344/IOC.ML.11.2.

(606) *Mammal Diversity Database (Version 1.5)*, 2021. DOI: 10.5281/zenodo.4139818.

(607) Royal Botanic Gardens; Kew. *State of the World's Plants and Fungi, 2020*; Royal Botanic Gardens, Kew, 2020.

(608) Lowenstein, J. M. Species-Specific Proteins in Fossils. *Naturwissenschaften* **1980**, *67*, 343−346.