ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/263584013

Protein Molecular Data from Ancient (> 1 million years old) Fossil Material: Pitfalls, Possibilities and Grand Challenges

ARTICLE *in* ANALYTICAL CHEMISTRY · JULY 2014

Impact Factor: 5.64 · DOI: 10.1021/ac500803w · Source: PubMed

READS

381

3 AUTHORS, INCLUDING:

Mary H Schweitzer

North Carolina State University

74 PUBLICATIONS **1,254** CITATIONS

SEE PROFILE

Elena R. Schroeter

North Carolina State University

7 PUBLICATIONS **3** CITATIONS

SEE PROFILE

Available from: Mary H Schweitzer
Retrieved on: 01 February 2016

analytical chemistry

Perspective

pubs.acs.org/ac

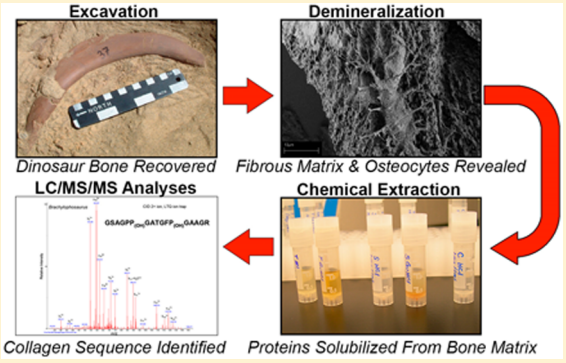
Protein Molecular Data from Ancient (>1 million years old) Fossil Material: Pitfalls, Possibilities and Grand Challenges

Mary Higby Schweitzer,*[,†,‡] Elena R. Schroeter,*[,†] and Michael B. Goshe*[,§]

[†]Marine, Earth and Atmospheric Sciences, [§]Molecular and Structural Biochemistry, North Carolina State University, Raleigh, North Carolina 27695, United States

[‡]North Carolina Museum of Natural Sciences, Raleigh, North Carolina 27601, United States

ABSTRACT: Advances in resolution and sensitivity of analytical techniques have provided novel applications, including the analyses of fossil material. However, the recovery of original proteinaceous components from very old fossil samples (defined as >1 million years (1 Ma) from previously named limits in the literature) is far from trivial. Here, we discuss the challenges to recovery of proteinaceous components from fossils, and the need for new sample preparation techniques, analytical methods, and bioinformatics to optimize and fully utilize the great potential of information locked in the fossil record. We present evidence for survival of original components across geological time, and discuss the potential benefits of recovery, analyses, and interpretation of fossil materials older than 1 Ma, both within and outside of the fields of evolutionary biology.



Ever since the early work of Abelson[1−4] hinted at the possibility that very old fossils may still contain some original organic content, there have been periodic efforts to optimize the search for informative biomolecules retained in the matrices of fossil remains. Early efforts were plagued by technological limitations, lack of specificity and controls, and inadequate databases against which to compare recovered sequences.[5−18] The advent of increasingly sensitive and higher resolution methods in mass spectrometry (MS), coupled with a greatly expanded proteomic database, have increased the possibility of recovering evolutionarily significant molecular information from many fossils, even those from the Mesozoic (~250−66 Ma).[19−27] However, to take full advantage of these advances, a systematic analytical approach incorporating new sample preparation, acquisition, and data analysis methods must be developed to optimize the detection of informative biomolecules from the ancient past.

DNA codes all structural RNAs and proteins, which in turn determine organismal fitness. Protein sequences can be derived from corresponding DNA sequences, but the three-dimensional structure, post-translational modifications (PTMs) and intramolecular bonds of proteins are not directly apparent from DNA sequence. Yet *these* characteristics are what ultimately determine protein function; thus, these gene products are the ultimate targets of natural selection.

DNA and protein sequences are *conceptually* similar, but chemically distinct. When DNA is present and recoverable in fossils, it can be interpreted within the established framework of molecular evolution (e.g. ref 28 and references therein); however, factors influencing its preservation over geological time are not well understood. Proteins have chemical diversity far exceeding that of DNA, and this diversity confers differential preservation. Thus, while DNA remains the "holy grail" of ancient molecular research for the majority of those working to recover molecular information from fossils, we have chosen to focus on the preservation of proteins and/or peptides, because: (1) Proteins can potentially outlast DNA, pushing back the time frame for recovery of phylogenetically informative molecules.[29,30] (2) Protein detection does not rely on artificial amplification, so there is less susceptibility to contamination.[31,32] (3) A broad spectrum of proteins can be characterized from small amounts of sample material, requiring less destruction and providing more repeatability. (4) Antibodies to specific protein or peptide components can be used to identify and separate proteins from a heterogeneous mixture and do not require complete proteins for binding, making them ideal tools for studying degraded or fragmented ancient proteins (e.g. refs 33,34). (5) Antibody−antigen reactivity provides an inherent cross-check of other fossil-derived data, including sequence data, when it can be shown that antibodies to extant purified proteins bind to fossil components, and antibodies raised against fossil compounds bind to purified extant proteins. Thus, if sequences for a particular protein are obtained from fossil extracts, these can be independently confirmed with specific antibodies; conversely, antibody binding can direct database searches for identified proteins.

Received: March 1, 2014
Accepted: June 20, 2014

 © XXXX American Chemical Society dx.doi.org/10.1021/ac500803w | *Anal. Chem.* XXXX, XXX, XXX−XXX

MS ANALYSIS: OLD VS ANCIENT SAMPLES

The analysis of proteins by mass spectrometry (MS) can be achieved by a wide variety of instrument platforms and acquisition methods (refs 35−38 and references therein). The two major approaches involve top-down or bottom-up analysis. For the top-down approach, samples containing intact proteins are directly analyzed, whereas the bottom-up approach analyzes peptides resulting from proteolytically digested proteins. When both methods are applied to a given sample, they can provide complementary information to more comprehensively identify a protein and its PTMs. Both top-down and bottom-up methods use liquid chromatography−tandem mass spectrometry (LC/MS/MS) to separate analytes by reversed-phase liquid chromatography for MS analyses. Hybrid mass spectrometers containing quadrupole collision cells or linear ion traps coupled to time-of-flight, orbitrap, or Fourier transform ion cyctotron resonance mass analyzers are able to fragment analytes of interest to generate product ion spectra of precursor ions (ionized proteins or peptides) that provide structural information regarding amino acid residue sequence and PTMs. The LC/MS/MS systems in wide use today are typically capable of high-resolution separations, high mass measurement accuracy, and high resolving power, which are required to generate more accurate mass measurements to characterize peptide and protein precursor ions and their corresponding product ions, which, in turn, can aid in spectral interpretation when compared to instruments of lower mass measurement accuracy and resolving power.

Mass spectrometry has been applied to fossils, but it has primarily been confined to samples that we would consider old but not ancient (i.e., <1 Ma). Some examples (not an exhaustive list) of utilizing bottom-up proteomic approaches on old (archeological) samples include analyzing egg white in pigments from 500 to 600 year old works of art for dating purposes using quadrupole-time-of-flight (Q-TOF) LC/MS/MS,[39] identifying enamel protein in mummy teeth (1100 years old) using matrix-assisted laser desorption ionization (MALDI)-TOF/TOF,[40] amino acid analyses of ancient grape seeds (∼1300 years old) by pyrolysis gas chromatography−mass spectrometry (GC/MS) and MALDI-TOF and LC/MS/MS for proteome analysis,[41] analyzing animal hair from the clothing of "Oetzi" (a Tyrolean mummy >5300 years old) by peptide mass finger printing and MS/MS analysis using MALDI-TOF/TOF,[42,43] sequencing proteins from a mammoth femur (43,000 years old) by LC/MS/MS using an orbitrap instrument,[44] and sequencing bone collagen (archeological samples with a thermal age of >100 $ka_{collagen@10\ °C}$) using MALDI-TOF MS and MS/MS analysis.[45] Although these samples range from 500 to 100,000 years old, they share, to a varying degree, common challenges: the extent of protein preservation within the sample, the necessity of comparing modern control samples to archeological samples, minimizing contamination from the surrounding environment, and the lack of suitable genome databases (and knowledge regarding novel PTMs) required for more comprehensive analyses to fully characterize protein components.

For our work, we have defined "ancient" as >1 Ma, consistent with current, but flawed, hypotheses of limits to DNA longevity (e.g., refs 46−52). Although arbitrary, this age represents a conceptual threshold past which analytical challenges become increasingly magnified. These hypothesized limits to preservation are based on extrapolation from degradation under unrealistically harsh chemical conditions,[46] chemical models that overlook the stabilizing influence of close mineral association on molecules,[53−55] or extrapolation of data from a limited set of fossils,[56] and cause reports of molecular recovery from fossils older than 100,000 to be greeted with skepticism,[46,47,49−51,57] (but see refs 30,58).

Conversely, the limits of protein survival are not known, but exceed that of DNA, as proteins have been recovered and validated from significantly older material[20,21,23−27,59,60] than authenticated DNA. Protein sequences derived from fossils >1 Ma have been used in phylogenetic reconstruction[23,61] and to address questions of molecular function.[60]

THE CHALLENGES

The dilemma in working with ancient molecules is, of course, that they are ancient—and therefore highly modified in ways that are not accounted for in existing analytical techniques or comparative databases. There is no doubt that the decrease in abundance and concentration of endogenous protein components present in a sample and the relative increase in accompanying contaminants (contained within the sample and/or surrounding environment) make analyses of older fossils exceedingly difficult. Molecular signals arising from older fossils could be microbial in origin or could result from unpredictable reactions with endogenous or exogenous compounds (e.g., Maillard reactions,[62] sulfurization,[63] or through other pathways[64,65]), as well as artifacts due to sample aging and/or preparation (e.g., deamidation[66−68]) all of which could impede protein/peptide ionization for MS analysis and species assignment of identified sequences. Additionally, fossils older than 1 Ma may be evolutionarily distinct, even from their close extant relatives, and current search algorithms may not recognize preserved sequences that differ significantly from those residing in existing databases.

Sample Preparation. It is apparent that rational, systematic, and robust sample preparation methods need to be established for routine use with ancient samples, facilitating comparison of results across various sample types and laboratories. This includes rigorous comparison, and perhaps standardization, of methods used to extract proteins from fossil tissues for analyses. Extraction protocols for fossil samples have used varying concentrations of different reagents for the demineralization and solubilization of proteins from ancient bone, including (but not limited to) sodium ethylenediaminetetraacetic acid (EDTA),[22,44,69,70] ammonium EDTA,[71] hydrochloric acid (HCl),[30,72] guanidine hydrochloride (GuHCl),[22,24] and ammonium bicarbonate.[44,72] However, these solutions, and others previously tested on bone proteins (e.g., urea/thiourea, SDS) have been shown to vary in their efficacy by enzyme-linked immunosorbent assay (ELISA) and SDS-PAGE,[73] and it is currently unknown if there is a disparity between the quantity, quality, and type of proteins they extract for mass spectrometry analyses. Indeed, recent mass spectrometric comparison of insect proteins solubilized by different reagents showed that each solution yielded not a complete proteome, but a unique subset of proteins comprising it (i.e., an "extractome").[74]

In any proteomic experiment, the most abundant proteins are detected; however, identification of low-abundant protein components usually requires another sample preparation step that employs enrichment or depletion methods, or orthogonal chromatographic fractionation prior to online reversed-phase separation. This type of approach was recently applied to

dx.doi.org/10.1021/ac500803w | Anal. Chem. XXXX, XXX, XXX−XXX

Table 1. Molecular Evidence for Persistence of Proteins in Ancient Vertebrate Material Generated in Our Labs and Published As Cited

protein	sequence[c]	*in situ* antibody reactivity	specimens
collagen α1t1	GATGAP*GIAGAP*GFP*GAR	+	*T. rex*,[20−22,59] *B. canadensis*[24]
collagen α1t1	GSAGPP*GATGFP*GAAGR	+	*T. rex*,[20−22,59] *B. canadensis*[24]
collagen α1t1	GVQGPP*GPQGPR	+	*T. rex*,[20−22,59] *B. canadensis*[24]
collagen α1t1	GVVGLP*GQR	+	*T. rex*,[20−22,59]
collagen α1t1#[a]	GAPGPQGPSGAP*GPK	+	*T. rex*,[20−22,59]
collagen α2t1	GLPGESGAVGPAGP*IGSR	+	*T. rex*,[20−22,59] *B. canadensis*[24]
collagen α1t1	GLTGPIGPP*GPAGAP* GDKGEAGPSGPPGPTGAR	+	*B. canadensis*[24]
collagen α1t1	GETGPAGPAGPP*GPAGAR	+	*B. canadensis*[24]
collagen α1t1	GPSGPQGPSGAP*GPK	+	*B. canadensis*[24]
collagen α2t1	GSN(deam)GEP*GSAGPP*GPAGLR	+	*B. canadensis*[24]
osteocalcin		+	*T. rex*,[22] *B. canadensis*[24]
actin	AGFAGDDAPR	+	*T. rex*,[25] *B. canadensis*[25]
actin	AVFPSIVGR	+	*B. canadensis*[25]
tubulin α-1A	QLFHPEQLITGK	+	*T. rex*,[25]
tubulin α-1A	EIIDLVLDR	+	*T. rex*,[25] *B. canadensis*[25]
histone H4	VFLENVIR	+	*B. canadensis*[25]
histone H4	TVTAMDVVYALK	+	*B. canadensis*[25]
histone H4	TLYGFGG	+	*B. canadensis*[25]
histone H4	ISGLIYEETR	+	*T. rex*,[25] *B. canadensis*[25]
histone H4	DAVTYTEHAK	+	*B. canadensis*[25]
hemoglobin b-A[b 59]	VNVADC(cam)GAEALAR	+	*T. rex*[59]
hemoglobin b-A[59]	LSDLHAQK	+	*T. rex*[59]
hemoglobin		+	*B. canadensis*[24]
chicken phospho-endopeptidase (PHEX)[25,111,112]		+	*T. rex*,[25] *B. canadensis (mab)*[25]
DS DNA backbone (Millipore MAB030)[25]		+	*T. rex*,[25] *B. canadensis (mab)*[25]
β-keratin		+	*S. deserti*,[105] *R. ostromi*[104]

[a]Collagen α1t1#: This collagen sequence is unique to *T. rex* in current databases.[20,21,101] [b]Hemoglobin sequences obtained matched several birds in available databases, but this sequence had a ByOnic score of 782, almost twice that of the other sequence.[59] [c]The "*" refers to a hydroxyl modification; "deam" refers to deamidation; "cam" refers to carbamidomethylation.

ancient samples ranging from 6 ka to 1.5 Ma to characterize fossil bone proteomes.[30] Proteins were extracted from whole bone with bacterial collagenase to test the hypothesis that the selective enzymatic destruction of collagen, the most abundant protein in bone, would allow for the detection of less abundant noncollagenous proteins (NCPs). Collagenase extraction products were compared with those resulting from incubation with GuHCl on an orbitrap, using a bottom-up LC/MS/MS approach. The authors found that, in both modern and fossil bone, treatment with collagenase yielded extractomes of much *lower* complexity than those generated by extraction with GuHCl. Furthermore, in all but the modern bone, collagenase failed to remove collagens as the top scoring proteins in database searches and failed to improve detection of NCPs. This is most likely because collagenase cleaves collagens at the X-Gly bond in the frequently occurring Pro-X-Gly-Pro sequence,[75,76] but does not *remove* cleaved product from the sample; thus these fragments remain in the sample, contributing to the masking of less abundant NCPs, and lead to the consequent persistence of detectable collagens in ancient samples—particularly if incubation time was insufficient to fully digest all the collagen present. Alternatively, collagenase may be a poor reagent for the initial solubilization of NCPs from bone matrix. Enzymatic treatment of extraction products from GuHCl incubation (or another extraction buffer) could yield better results than generated by either treatment alone. Other approaches, such as immunoprecipitation or reverse affinity purification, alone or in conjunction with collagenase and other treatments should be explored as avenues to exclude more abundant proteins, thus allowing recovery of less abundant, but potentially more phylogenetically informative portions of ancient proteomes.

Interestingly, although collagenase digestion did not show higher NCP recovery over GuHCl, the authors were able to recover α-2-HS-glycoprotein (A2HSG) in the GuHCl extracts of ancient samples.[30] They suggested that A2HSG could be useful as a marker for mapping species differentiation, because this blood and serum protein is incorporated into the matrix during bone formation and is hypothesized to be involved in the initial steps of bone mineralization. In addition, because unlike albumin, A2HSG is not a common laboratory reagent/standard, the likelihood that its detection arises from contamination is reduced. Although this preparation method was not effective in removing dominant proteins to increase detection of less abundant ones, this study shows that increased sequence detection through alternative sample preparation methods must be further explored in the field of paleoproteomics.

Protein/Peptide Sequencing and PTM Analyses. Advanced database search algorithms (e.g., ByOnic) are improving the accuracy and identification of sequences recovered from fossils. An example of this approach for ancient samples is illustrated by the sequence data presented in Table 1. The experiments producing the cited sequences implemented a

dx.doi.org/10.1021/ac500803w | *Anal. Chem.* XXXX, XXX, XXX−XXX

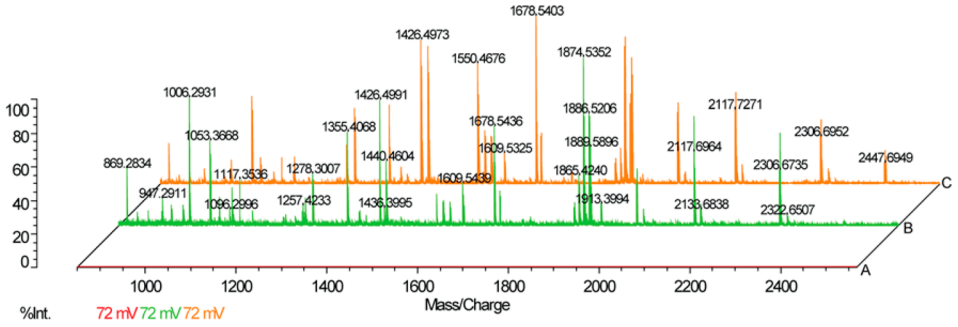


Figure 1. MALDI-TOF-MS Spectra of *Brachylophosaurus canadensis* (MOR 2598) extracts spiked with 250 fmol of rabbit phosphorylase b. (A) presample treatment, (B) postsample treatment with chelator, and (C) phosphorylase b standard. The ion suppressant present in the *B. canadensis* sample (A) completely prevented detection of the phosphorylase b internal standard. After sample processing, detection of the internal standard in the *B. canadensis* sample (B) matched nearly all of the *m*/*z* values and intensities of the intact internal standard (C). Additional peaks in (B) not present in (C) are unique analytes obtained from the *B. canadensis* sample.

bottom-up proteomics analysis using LC/MS/MS. The raw data were analyzed and sequences confirmed with various search algorithms,[23,24,59] and novel sequence and PTM information contained in several dinosaur bone samples were identified.

In addition, the use of *de novo* sequencing algorithms rather than database searching may yield novel amino acid sequence data and characterize PTMs that may both contribute to protein preservation and reflect original species diversity. Because sequences are derived directly from the measured MS/MS data in *de novo* sequencing, without database-matching predicted spectra,[77] the lack of known genomes/proteomes for phylogenetically similar extant taxa is not a hindrance for analyzing extinct taxa.

Deamidation of asparigine residues, and to a lesser extent, glutamine residues, is a well-known age-related modification in ancient proteins. A recent paper used a combination of MALDI-TOF-MS and Q-TOF LC/MS/MS analyses to characterize the deamidation of α-keratin peptides in ancient wool textile samples from different locations (United Kingdom and Iceland) ranging from the ninth to 16th century[78] and exemplifies the challenges encountered in analyzing and interpreting chemical modifications in archeological samples. A marker set of eight peptides ranging from two to five deamidation sites was shown to be suitable for characterizing protein deamidation in sheep wool. It was determined that three of these peptides could not be used to determine deamidation rates due to a lack of detection of deamidation in archeological samples that may be the result of peptides being (1) "lost" from the samples; (2) rapidly deamidated; or (3) modified during sample preparation or processing (e.g., dyeing) of the material as part of their original use. Interestingly, measured deamidation sites did not necessarily correlate with their nonenzymatic rates based on primary structure (i.e., sequence), indicating that either higher-order structural elements of the α-keratin or aspects of the depositional environment altered deamidation rates, as was the case for acidic soils which promoted wool preservation but also favored deamidation. In fact, the deamidation rates of the ancient samples correlated poorly to their age, whereas the rates of aged wool in a controlled environment displayed time-dependent deamidation. As discussed by the authors, understanding the chemical processing of the wool and its exposure to environmental factors needs to be considered to better interpret deamidation profiles. Because deamidation can also occur during preparation of samples for analyses,[68] other experimental steps and controls may be required to differentiate archeological-based deamidation from those occurring from sample preparation, such as employing heavy water ($H_2{}^{18}O$) during sample processing to account for preparation-induced deamidation events.[79]

Although processes such as deamidation may be accounted for in archeological protein analysis, many of the acquired product ion spectra generated for ancient samples are unassigned, and may result from unknown pathways or mechanisms contributing to preservation; thus, greater attention to identifying PTMs is required. Although this is a time-consuming endeavor that requires additional analytical methods (e.g., nuclear magnetic resonance (NMR)) to validate PTMs in ancient samples, the detection of novel PTMs can be achieved. An excellent example is the identification of sulfilimine bond (−S═N−) cross-links that occur between hydroxylysine 211 and methionine 93 within collagen IV networks.[80] This PTM appears to have occurred at the time of divergence of sponge and cnidaria and was proposed by the authors to be an adaptive response to the mechanical stress encountered by these organisms. Combining these bottom-up proteomic approaches with top-down analysis and additional biochemical validation will establish a baseline of information from which future ancient samples can be interpreted. Additionally, the data can be used to transition to a more high throughput and targeted quantitative analysis, such as selected reaction monitoring (i.e., multiple reaction monitoring (MRM), a multiplexed SRM)[81,82] to permit the comparison of samples across time and species, thus providing more insights regarding species survival and differentiation.

However, the very processes that may contribute to preservation (e.g., cross-linking, exogenous and/or endogenous reactions, or novel PTMs) may also contribute to inhibition of analyses. The intra- and intermolecular cross-links that stabilize molecules like collagen and keratin may accumulate over time and become irreversible, yielding unpredictable results or ambiguous data. These cross-links may render molecules unavailable for trypsin digestion, or cause them to be fragmented in an atypical manner, resulting in peptide fragments that differ greatly from predicted patterns established for controlled enzymatic degradation of extant proteins.[64] Alternatively, they may incorporate exogenous components from geochemical or microbial influences that act on tissues over time. Because mass spectrometric database searches rely on predicted masses of precursor/product ions for sequence identification, such modifications can confound the detection of

dx.doi.org/10.1021/ac500803w | *Anal. Chem.* XXXX, XXX, XXX−XXX

ancient proteins that are present. For example, in lab-based experiments we have found that iron may play a role in the initial stabilization that results in preservation of soft tissue structures and the molecules that comprise them, but may also cause ion suppression (Figure 1[83]), and similarly, may interfere with enzymes used in DNA sequencing. This phenomenon may explain in part why we can demonstrate antibody binding to proteins *in situ*, but cannot yet identify them using mass spectrometry analyses, and why the detection of low concentrations of fossil biomolecules may be the ultimate analytical challenge.

Another difficulty is that interpretation of sequence data, whether from proteins or DNA, relies on bioinformatics tools, either database matching or *de novo* peptide sequencing. Applying these methods to data obtained from fossils poses significant problems. Efficient and accurate database searches require a sequence library from closely related extant taxa—which does not exist for most fossil samples. This is changing with the addition of whole genome sequencing of non-mammalian organisms, but mammals still greatly outnumber other organisms. Additionally, current bioinformatics search algorithms ensure that the most homologous or conserved peptides will be identified, but if identified, these matches can be claimed as evidence for contamination. Thus, proteins or peptides could be present in fossil material, but may not be recognized by current search algorithms if they differ significantly from those of living mammals, or may be discounted as contamination if too similar to existing entries. Conversely, while *de novo* sequencing is not so dependent on an existing database of peptide information, it does require high mass accuracy and high mass resolution MS/MS data to accurately interpret novel sequences.[77,84] Achieving such precise fragmentation data may be challenging in fossil samples, where cross-linking and other diagenetic pathways may interfere with efficient fragmentation (see above discussion), not to mention the difficulty of obtaining a pure sample from such compromised starting material. To overcome this obstacle, alternative fragmentation methods to collision-induced dissociation (CID) (e.g., electron transfer dissociation (ETD) or electron capture dissociation (ECD)), or combinations of these approaches, should be attempted to obtain a more complete fragmentation ladder, allowing for more accurate sequencing.[84,85] The efficiency/accuracy of automated data interpretation (e.g., no allowances for relative mass accuracy of fragment ions, see ref 84 for discussion), is lacking in much of the current *de novo* software, but future technological advancements in these programs will undoubtedly continue to make *de novo* sequencing a more viable alternative to database searching for extinct proteomes.

These analytical and bioinformatics approaches must account for both recognition of sequences not residing in existing databases, and the special characteristics/modifications of ancient biomolecules (e.g., damage/decay patterns) not anticipated by existing technology. Furthermore, differences between fossil sequences and extant counterparts may arise from evolutionary change, or from subsequent post-mortem or diagenetic reactions. Distinguishing between these is a significant challenge, particularly when databases are limited, and complicates *de novo* assembly of sequenced DNA or proteins when unknown modifications may be present, and because some of these modifications may contribute to ion suppression (see below) in mass spectrometry analyses or enzyme interference for DNA studies, recognition of endogenous molecules is further compromised.

In summary, to cope with low abundances, prevalent but unusual modifications, and the lack of protein databases for extinct taxa, several complementary sample preparation and MS approaches can be employed. Various proteases (i.e., trypsin, Glu-C, Arg-C) may be applied to fossil extracts to achieve higher protein coverage and peptides of optimal lengths for MS approaches, and alternative fragmentation methods will increase recovery of proteomic information retained in fossil materials. Bioinformatics approaches must be developed to include modification- and mutation-tolerant database search and *de novo* sequencing. Homology-guided *de novo* sequencing[86] has potential to expand recoverable sequence data from known or suspected proteins such as collagen from extinct organisms. Targeted peptide detection methods such as SRM/MRM provide greater sensitivity and throughput. When antibody binding indicates the presence of particular proteins that may be masked under standard, untargeted MS techniques, these targeted methods may yield additional data.

Contamination. Finally, there is the issue of contamination. When dealing with a sample highly enriched with proteins and containing few contaminating proteins (usually keratin and trypsin, which are readily accounted for during database searching), as is common for samples derived from extant organisms, the sample proteins are easily identified in most MS protocols; conversely, the exogenous contaminating proteins may or may not be detected on the basis of their relative abundance to the endogenous proteins present in the sample. On the other hand, when very low abundance, highly modified, and fragmented proteins are contaminated with exogenous proteins, the contaminants are preferentially identified because of their high relative abundance compared to that of the endogenous protein components. If the contaminants arise from a modern sample closely related to the ancient one, perhaps even coanalyzed as a positive control, it may well be impossible to distinguish endogenous ancient signals from those arising from contamination. While adequate use of controls in well-designed experiments can identify or account for contamination, if a modern contaminant has gone through rounds of sterilization (UV may trigger unusual cross-links, bleach may oxidize residues), it may still be recognizable but chemically modified in a manner similar to truly ancient protein components.

For these reasons, it is *critical* that multiple negative controls are used throughout the analytical process to account for any exogenous contamination that may have been introduced during the processing, handling, excavation, and diagenetic history of a sample. In our analyses, we routinely use a series of controls that each address a potential contaminant source. To exclude the entombing environment of a fossil as a source of exogenous proteins (e.g., from microbial organisms over its burial history, or handling during excavation), we collect sediments immediately adjacent to (and sometimes infilling) our fossils, and conduct all analyses on these sediments in tandem with their fossil counterparts, using solutions drawn from the same stock. "Blank" samples, consisting of only buffer solutions and reagents used for the assays, are run as a second negative control to detect any contaminants that might be present in the lab or those to which the samples might have been exposed to during experimental procedures. Each immunological assay is performed in conjunction with specificity tests to account for spurious binding. These include

dx.doi.org/10.1021/ac500803w | *Anal. Chem.* XXXX, XXX, XXX−XXX

the following: (1) inhibition (also called "preadsorption") assays, in which the paratopes present in an antibody are blocked with its purported immunogen before incubation with the sample, allowing the identification of additional paratopes contained in our polyclonal antibodies to molecules other than the target protein;[87−89] (2) digestion assays, in which the target protein is selectively destroyed (either partially or completely) before incubation with antibodies raised against it, to assess whether antigen−antibody complex formation is directly affected by the specific digestion of the target protein; (3) irrelevant antibody assays, in which samples are incubated with antibodies raised against proteins that are not typically found in the sample tissue (e.g., mammalian antitestosterone antibodies with dinosaur bone tissue) to confirm that the reactions to sample material are not indiscriminate.[89] These are in addition to standard "secondary-only" assays included in every trial of every experiment, where a subset of samples are incubated with antibody buffer (lacking primary antibody) during the primary antibody step, then treated to secondary antibody and all other steps for signal detection to determine if the secondary may be binding nonspecifically to molecules in the tissue, and not exclusively to bound primary antigen−antibody complexes.

While methodological design can account for contamination and control for false positives, it is also imperative to decrease the risk of contamination from modern sources as much as possible. Testing "modern controls," (extant taxa phylogenetically related to the fossil taxon), is necessary to assess the validity and efficiency of any given assay for detecting a target protein. For example, if chicken-specific anticollagen I antibodies are to be used against fossil dinosaur tissue (reasonable, as chickens and dinosaurs are phylogenetically close[90]), they should first be confirmed to bind with collagen I in chicken bone; if such assays on extant bone give poor or negative results, then any immunoreactivity observed in the tissue of ancient, distantly related taxa should be suspect. Alternatively, if no binding is observed in either modern or ancient samples, it may indicate procedural error rather than no recoverable epitopes in ancient samples.

However, testing on these extant samples could potentially contaminate ancient samples, if they are conducted in the same laboratory space. To minimize exposure of fossil samples to modern proteins, we conduct all analyses of ancient material in rooms separate from those of extant controls, where duplicate equipment is dedicated for this purpose. By completely separating storage and experimentation of ancient samples (and their negative controls) and positive control assays in space, we ensure that fossil material never comes into contact with solutions, instruments, containers, or analytical equipment used for extant tissues, greatly reducing the opportunity for contamination with modern extracellular matrix proteins. Additionally, aseptic techniques are used for all sample handling. Fossil samples are only processed and prepared wearing nitrile (nonlatex) gloves and full-length lab coats, shoe covers, and hairnets, and these are required when entering space dedicated to fossil analyses. Tools and surfaces are sterilized with 95% ethanol followed by 10% bleach (and, when possible, autoclaving).

However ultimately, the most compelling evidence for endogeneity of proteins in ancient samples is obtaining sequence data that aligns with phylogenetic predictions.[52,91] This is the end goal of paleoproteomics, and the reason compelling all other advancements. Only by sequence comparison can we begin to fully utilize recovered protein sequences to better understand rate, direction and pattern of molecular evolution.

CASE STUDY

The preservation of endogenous molecules within the bone of dinosaurs is controversial, because chemical models and benchtop experiments that employ unrealistically harsh conditions as proxies for time and degradation indicate that biomolecules have a limited "shelf-life"[46,55] (but see refs 53,54,92−94) considerably shorter than the ∼65 Ma since the last nonavian dinosaur walked the earth. Deamidation, strand breaks, backbone cleavage, post mortem modifications, abnormal cross-links, oxidation, and base or amino acid alteration and/or substitution[46,51,91,95] will render molecular sequences uninformative, misleading, or difficult to attain. However, it is recognized that environmental conditions, particularly water/humidity and thermal history, are more important than age in determining survivability of original molecules in fossil matrices,[30,55,96] and the arbitrary one million year limit for DNA is being challenged.[58] Nevertheless, claims for older endogenous molecules are met with skepticism, and very few workers attempt analyses on fossils older than one million years.

In addition, if the community finds it unacceptable that small fragments of highly altered, but original *molecules* may persist over geological time, much greater skepticism is brought to bear over the possibility that tissues and/or cells themselves could persist, and still retain at least some of their original features of transparency, flexibility/elasticity, and response to specific antibodies. Yet that is what we observed and first reported from an exceptionally well preserved *Tyrannosaurus rex* in 2005[97] (Figure 2).

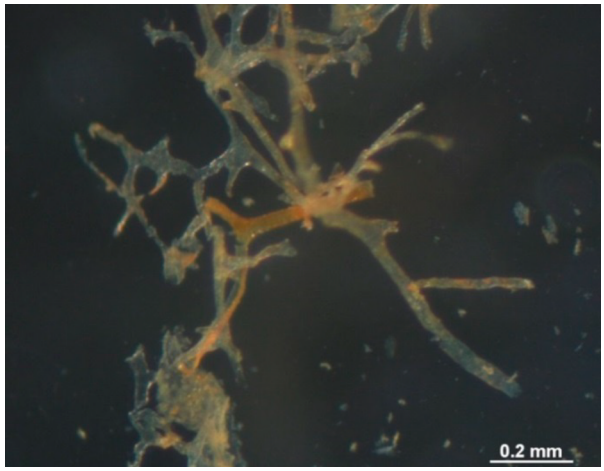


Figure 2. Cluster of transparent, flexible vessels recovered from *Tyrannosaurus rex* (MOR 1125) demineralized bone.

We followed this report with a chemical characterization of the demineralized, fibrous and flexible bone matrix,[20−22] that in living organisms, consists primarily of fibrillar, cross-linked, mineralized collagen I.[98] In this *T. rex*, and a second well preserved hadrosaurid dinosaur, *Brachylophosaurus canadensis*,[24] we showed that when either chemical extracts of whole bone powder or sections of demineralized matrix were exposed to antibodies raised against avian collagen I, reactivity followed the pattern of modern tissues, although with less intensity, and we produced MS sequence data consistent with collagen. We were

dx.doi.org/10.1021/ac500803w | *Anal. Chem.* XXXX, XXX, XXX−XXX

also able to demonstrate, using *in situ* TOF-secondary ion mass spectrometry (SIMS), that the collagen-specific post-translationally modified hydroxyproline and hydroxylysine residues were present in these tissues, and the characteristic banding pattern was evident using atomic force microscopy (AFM), although was not visible in transmission electron microscopy (TEM).[20−22,24] We then conducted a survey of extant and fossil bone from the present to Triassic specimens, and observed at least three out of the four original components (fibrous matrix, vessels, intravascular material, and osteocytes) in about half of the specimens, independent of age, taxon, or depositional setting.[99] The immunological and MS data, while controversial,[20,52,100,101] are consistent with the presence of original molecules and tissues in the fossils examined to date, and our hypothesis that these organic materials are original to these Mesozoic fossils and produced by the once-living animals is supported, despite conventional wisdom that claims a temporal limit for molecules and tissues.

Our laboratories seek to understand the persistence of endogenous molecules across deep time. Because our preliminary data from microscopic, immunological, and chemical assays, coupled with mass spectrometry analyses on extracts of fossil specimens (Table 1) supports the preservation of original components, the data suggest that proxies currently put forth to explain degradation over time are not accurate. So how DO labile tissues, and the molecules comprising them, enter the rock record? We need new models.

To persist in the rock record across geologic time, stabilization of organics must outpace decay.[102,103] Many factors may contribute to molecular preservation over deep time. For example, the molecular configuration of collagen itself may contribute to its selective preservation. When we mapped recovered sequences from two dinosaurs onto a molecular and functional model of extant collagen, it became apparent that those peptides remaining in dinosaur bone mapped to highly conserved, internal regions of the molecule, where they were well-protected from outside influences.[60]

After bones and teeth, epidermally derived, keratin-containing materials are most commonly represented in the vertebrate fossil record. The molecular structure of keratin may contribute to preservation in a manner similar to collagen, by incorporating helical structure, multiple cross-links, and hydrophobic residues that exclude water.[104,105] Three-dimensional chemical conformation may be key for the *initial* stabilization of molecules and the tissues they comprise,[60] especially when closely associated with mineral, as is easily accomplished in bony tissues.

Recently, we showed that iron may be a second factor in the preservation of cells, tissues and molecules across geological time.[83] We proposed that as biologically active, highly reactive Fe^{2+} is released from degrading materials containing iron (e.g., red blood cells, muscle tissues), it would undergo Fenton-type reactions, generating hydroxyl radicals in the process. Two results would follow: (1) Fe^{3+} would precipitate on organic surfaces as iron nanoparticles, which could then convert to goethite; and (2) hydroxyl radicals would form cross-links in molecules comprising the observed cells and tissues, stabilizing the molecule and acting to "fix" the tissues against degradation by blocking active sites targeted by enzymes of degradation. Both observational and experimental data support this hypothesis and provide a naturally occurring mechanism for initial stabilization, and eventually preservation, of soft tissue structures and the molecules that comprise them.[83]

THE POSSIBILITIES

So why try to recover ancient molecules from "fossil" organisms? It is a time intensive, extremely expensive and high-risk endeavor. Is it worth the effort? Because much depends on our ability to predict and prepare for future events, we *must* look to the past, for that is where the data on which to base these predictions reside. Organisms in the past adapted to face challenges that remain with us today, including limited (and limiting) resources, disease, environmental change, and natural disasters. While genetic coping strategies are being eroded in extant organisms through loss of diversity, knowledge of protein and/or DNA sequences from extinct organisms can shed light on *ancient* diversity. These data can be used to address current and future challenges facing life on our planet.

Experiments elucidating molecular characteristics of preserved soft tissues in ancient fossils can provide new insights into the giants that once ruled this world and the environments that allowed their success, as well as a new appreciation of the challenges and organismal adaptive responses to these challenges. For example, because dinosaurs share certain traits with living reptiles,[90] and living reptiles are ectothermic (cold-blooded), it was assumed that dinosaurs were also 'good reptiles.' However, morphological and biomechanical evidence supports an elevated metabolism in most dinosaurs, over that exemplified by extant ectotherms (e.g., ref 106 and references therein). Did nonavian dinosaurs achieve full endothermy, as exemplified by living birds and mammals, or was it a physiological strategy unique to this lineage, more efficient and less expensive in times of more equable global climates? We only have two living end points on the high end of the physiological continuum—did dinosaurs represent a third? Molecular data from fossil remains may shed light on the acquisition of elevated metabolic rates in these organisms, and may reveal other molecular traits that contributed to resilience in the face of past environmental changes, traits that ultimately produced our modern environments and the organisms which inhabit them.

Furthermore, by characterizing depositional/burial environments of fossils containing biomolecules using geochemical and sedimentological tools, we may discover environmental parameters that favor molecular preservation, and then target those environments to recover fossils for future molecular studies. Coupling this environmental information to the molecular biology of ancient organisms may also shed light on very real issues we face today. For example, such studies may elucidate molecular factors allowing organisms to adapt to varying CO_2 levels.

Collecting molecular data from extinct organisms and comparing them with data from living analogues will shed light on the rate and direction of evolutionary change at the molecular level. Among other outcomes, these data will be useful in validating molecular clocks and in independently testing existing hypotheses regarding evolutionary modes. Additionally, developing new bioinformatics approaches to analyze fossil data has immediate application beyond molecular paleontology. It has the potential to allow us to predict ancestral sequences, incorporate a mixture of proteomic and genomic data into single analyses, and estimate population genetic parameters by using data sets sampled from different time points that incorporate fossil data.

More broadly, characterization of ancient DNA or proteins may reveal a chemical basis for biomolecular preservation that

dx.doi.org/10.1021/ac500803w | *Anal. Chem.* XXXX, XXX, XXX−XXX

will increase our understanding of these molecules. Identifying molecular mechanisms contributing to preservation, such as novel cross-linking between moieties, or incorporation of functional groups as a result of long duration chemical reactions may lead to the creation of novel biomaterials that will improve the longevity of organo-electronics or some medical devices. Identifying regions of molecules that are particularly resistant to degradation and destruction may even improve the targeting and efficiency of pharmaceuticals.

Existing models of degradation propose that original materials have a definite temporal limit, but our research shows that iron (and possibly other metals) plays a major role in the early stabilization and long-range preservation of these materials, perhaps by generating radicals that participate in cross-linking reactions, thus acting as a fixative.[83] We have also shown that we can *remove* iron from these ancient tissues, improving our ability to recover original molecular fingerprints. Because iron has also been shown to contribute to many human diseases, including Alzheimer's Disease,[107] follow-up research on the role of iron in preservation may contribute to the development of novel treatments.

Other research shows that within a single organism, blood vessels exhibit "molecular heterogeneity", that is, different molecular characteristics depending upon the tissues they infiltrate, and their function.[108] Blood vessels derived from bone may differ fundamentally from vessels elsewhere. We suggest that bone-derived blood vessels express epitopes that differ from those of other vessels, and these epitopes may facilitate the deposition of mineral in forming bone. Indeed, during osteogenesis, vessels invade cartilaginous precursors, and bone deposits secondarily around them (ref 109 and references therein). This affinity for bone mineral may confer unexpected longevity to bone-derived vessels, and be part of the reason these tissues persist in fossil bone—but may also factor into the deposition of plaque (essentially bone, in mature form) on the *inside* of vessels in atherosclerosis. The search to explain the preservation of vessels in bone may shed light on aspects of cardiac/vascular disease in humans.

A more subtle benefit to the efforts to recover and characterize materials from ancient organisms, including dinosaurs, is perhaps equally important. Scientific literacy is declining in America at an alarming rate, yet the public fascination with dinosaurs is as high as at any point in history. It is for this reason that many museums and nonprofit organizations have utilized paleontology as a platform to design outreach programs that increase public interest in science, and this approach has been successful in engaging youth currently under-represented in scientific fields.[110] While paleontology provides an effective means to start a conversation with the public about science, the topics that these conversations cover may have a narrow focus if only traditional paleontological methods are considered; some of the keystone concepts of science, such as experimental design, positive and negative controls, reproducibility of results, and statistical evaluation of data, do not lend themselves easily to activities that focus only on the discovery and description of fossils. Molecular paleontology offers a natural dovetail between the excitement of studying "ancient beasts" and lab experimentation, and can be used as a framework to explore the scientific method, critical reasoning, experimental design, data analyses, and technological advancement, significantly increasing the breadth of science to which participants "hooked" by dinosaurs may become exposed.

SUMMARY

The investigation into molecular preservation in very old fossils is risky, expensive, and beset with many challenges and seemingly insurmountable obstacles. We believe the potential return is even greater than these obstacles. Pushing the limits of technologies that are required by this endeavor will benefit many disciplines and will allow us to ask questions thought unanswerable only a few years ago. A focused approach, employing the tools of analytical chemistry at all levels, will allow us to attain this goal more rapidly, but will also require a clear delineation of analytical standards and criteria for acceptance of endogeneity for any ancient material. Efforts to *fully characterize* ancient organic material, employing multiple analytical approaches across disciplines, will push the boundaries of all science. Much like the effort to put a man on the moon resulted in unintended consequences like transistor radios and Tang, it is possible that understanding molecular diagenesis, and improving the tools to detect it, will improve our lives as well, perhaps resulting in a better understanding of human disease and the coevolution of pathogens and hosts, and long-range predictions of responses to global change.

AUTHOR INFORMATION

Corresponding Authors

*M.H.S.: E-mail: Schweitzer@ncsu.edu. Telephone: 919-515-7838.

*E.R.S.: E-mail: easchroeter@ncsu.edu.

*M.B.G.: E-mail: mbgoshe@ncsu.edu. Telephone: 919-513-7740.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We would like to thank the reviewers, whose thoughtful comments greatly improved this manuscript. We also thank Wenxia Zheng, whose efforts in sample preparation and experimental design underlie much of this work; Tim Cleland, for comments and experimental design; Kevin Blackburn for sample preparation and MS analyses, Paul Thomas for endless hours of instruction and input, Neil Kelleher, John Yates, Marshall Bern, and others willing to take a chance on such an outlandish concept as analyses of multimillion year old remains, and the colleagues who have worked with us over the years to help demonstrate the feasibility of molecular analyses of fossils, as well as the critics of those published works, for their comments and skepticism have made the science (and the scientists) stronger for their diligence. Funding that allowed the experiments underlying this paper to be done include NSF INSPIRE (to M.H.S., M.B.G. and E.R.S.) and the David and Lucile Packard Foundation.

REFERENCES

(1) Abelson, P. H. *Science* **1954**, *119* (3096), 576.

(2) Abelson, P. H., Paleobiochemistry: organic constituents of fossils. In *Carnegie Institution of Washington, Yearbook 54*; Carnegie Institution of Washington: Washington, D. C., 1955; Vol. *54*, pp 107−109.

(3) Abelson, P. *Sci. Am.* **1956**, *195*, 83−92.

(4) Abelson, P. H. *Ann. N.Y. Acad. Sci.* **1957**, *69* (2), 276−285.

(5) Miller, M. F., II; Wyckoff, R. W. G. *Proc. Natl. Acad. Sci. U.S.A.* **1968**, *60* (1), 176−178.

(6) Matter, P. I.; Davidson, F. D.; Wyckoff, R. W. G. *Proc. Natl. Acad. Sci. U.S.A.* **1969**, *64*, 970−972.

(7) Towe, K. M.; Urbanek, A. *Nature* **1972**, *237*, 443−444.
(8) Weiner, S.; Lowenstam, H. A.; Hood, L. *Proc. Natl. Acad. Sci. U.S.A.* **1976**, *73* (8), 2541−2545.
(9) Weiner, S.; Lowenstam, H. A.; Taborek, B.; Hood, L. *Paleobiology* **1979**, *5* (2), 144−150.
(10) Lowenstein, J. M. *Philos. Trans. R. Soc., B* **1981**, *292* (1057), 143−149.
(11) Armstrong, W. G.; Halstead, L. B.; Reed, F. B.; Wood, L. *Philos. Trans. R. Soc., B* **1983**, *301* (1106), 301−343.
(12) Muyzer, G.; Westbroek, P.; De Vrind, J. P. M.; Tanke, J.; Vrijheid, T.; De Jong, E. W.; Bruning, J. W.; Wehmiller, J. F. *Org. Geochem.* **1984**, *6*, 847−855.
(13) Lowenstein, J. M. *Am. Sci.* **1985**, *73* (6), 541−547.
(14) Muyzer, G.; Westbroek, P. *Geochim. Cosmochim. Acta* **1989**, *53* (7), 1699−1702.
(15) Ambler, R. P.; Daniel, M. *Philos. Trans. R. Soc., B* **1991**, *333* (1268), 381−389.
(16) Collins, M. J.; Muyzer, G.; Westbroek, P.; Curry, G. B.; Sandberg, P. A.; Xu, S. J.; Quinn, R.; Mackinnon, D. *Geochim. Cosmochim. Acta* **1991**, *55* (8), 2253−2257.
(17) Gurley, L. R.; Valdez, J. G.; Spall, W. D.; Smith, B. F.; Gillette, D. D. *J. Protein Chem.* **1991**, *10* (1), 75−90.
(18) Logan, G. A.; Collins, M. J.; Eglinton, G., Preservation of Organic Biomolecules. In *Taphonomy: Releasing the Data Locked in the Fossil Record*, Allison, P. A.; Briggs, D. E. G., Eds.; Plenum Press: New York, 1991; Vol. *9*, pp 1−24.
(19) Schweitzer, M. H.; Hill, C. L.; Asara, J. M.; Lane, W. S.; Pincus, S. H. *J. Mol. Evol.* **2002**, *55*, 696−705.
(20) Asara, J. M.; Garavelli, J. S.; Slatter, D. A.; Schweitzer, M. H.; Freimark, L. M.; Phillips, M.; Cantley, L. C. *Science* **2007**, *317*, 1324−1325.
(21) Asara, J. M.; Schweitzer, M. H.; Phillips, M. P.; Freimark, L. M.; Cantley, L. C. *Science* **2007**, *316*, 280−285.
(22) Schweitzer, M. H.; Suo, Z.; Avci, R.; Asara, J. M.; Allen, M. A.; Teran Arce, F.; Horner, J. R. *Science* **2007**, *316*, 277−280.
(23) Organ, C. L.; Schweitzer, M. H.; Zheng, W.; Freimark, L. M.; Cantley, L. C.; Asara, J. M. *Science* **2008**, *320*, 499.
(24) Schweitzer, M. H.; Zheng, W.; Organ, C. L.; Avci, R.; Suo, Z.; Freimark, L. M.; Lebleu, V. S.; Duncan, M. B.; Vander Heiden, M. G.; Neveu, J. M.; Lane, W. S.; Cottrell, J. S.; Horner, J. R.; Cantley, L. C.; Kalluri, R.; Asara, J. M. *Science* **2009**, *324*, 626−629.
(25) Schweitzer, M. H.; Zheng, W.; Cleland, T. P.; Bern, M. *Bone* **2013**, *52*, 414−423.
(26) Buckley, M.; Larkin, N.; Collins, M. J. *Geochim. Cosmochim. Acta* **2011**, No. 75, 2007−2016.
(27) Lindgren, J.; Uvdal, P.; Engdahl, A.; Lee, A. H.; Alwmark, C.; Bergquist, K.-E.; Nilsson, E.; Edstrom, P.; Rasmussen, M.; Douglas, D.; Polcyn, M. J.; Jacobs, L. L. *PLoS One* **2011**, *6* (4), e19445.
(28) Shapiro, B.; Hofreiter, M. *Ancient DNA: Methods and Protocols*; Humana Press, Springer Science and Business Media: New York, NY, 2012; p 247.
(29) Buckley, M.; Anderung, C.; Penkman, K.; Raney, B. J.; Gotherstrom, A.; Thomas-Oates, J.; Collins, M. J. *J. Archaeol Sci.* **2008**, *35*, 1756−1764.
(30) Wadsworth, C.; Buckley, M. *Rapid Commun. Mass Spectrom.* **2014**, *28*, 605−615.
(31) Lubec, G.; Afjehi-Sadat, L. *Chem. Rev.* **2007**, *107*, 3568−3584.
(32) Corthals, A.; Koller, A.; Martin, D. W.; Rieger, R.; Chen, E. I.; Bernaski, M.; Recagno, G.; Davalos, L. M. *PLoS One* **2012**, *7* (7), e41244.
(33) Margolis, F. L. *Proc. Natl. Acad. Sci. U.S.A.* **1972**, *69* (5), 1221−1224.
(34) Anderson, D. J.; Blobel, G. *Methods Enzymol.* **1983**, *96*, 111−120.
(35) Beck, M.; Claassen, M.; Aebersold, R. *Curr. Opin. Biotechnol.* **2011**, *22* (1), 3−8.
(36) Kota, U.; Goshe, M. B. *Phytochemistry* **2011**, *72* (10), 1040−1060.
(37) Angel, T. E.; Aryal, U. K.; Hengel, S. M.; Baker, E. S.; Kelly, R. T.; Robinson, E. W.; Smith, R. D. *Chem. Soc. Rev.* **2012**, *41* (10), 3912−3928.
(38) Catherman, A. D.; Skinner, O. S.; Kelleher, N. L. *Biochem. Biophys. Res. Commun.* **2014**, *445* (4), 683−693.
(39) Gambino, M.; Cappitelli, F.; Catto, C.; Carpen, A.; Principi, P.; Ghezzi, L.; Bonaduce, I.; Galano, E.; Pucci, P.; Birolo, L.; Villa, F.; Forlani, F. *J. Biosci.* **2013**, *38* (2), 397−408.
(40) Porto, I. M.; Laure, H. J.; Tykot, R. H.; de Sousa, F. B.; Rosa, J. C.; Gerlach, R. F. *Eur. J. Oral Sci.* **2011**, *119* (Suppl. 1), 83−87.
(41) Cappellini, E.; Gilbert, M. T.; Geuna, F.; Fiorentino, G.; Hall, A.; Thomas-Oates, J.; Ashton, P. D.; Ashford, D. A.; Arthur, P.; Campos, P. F.; Kool, J.; Willerslev, E.; Collins, M. J. *Naturwissenschaften* **2010**, *97* (2), 205−217.
(42) Hollemeyer, K.; Altmeyer, W.; Heinzle, E.; Pitra, C. *Rapid Commun. Mass Spectrom.* **2008**, *22* (18), 2751−2767.
(43) Hollemeyer, K.; Altmeyer, W.; Heinzle, E.; Pitra, C. *Rapid Commun. Mass Spectrom.* **2012**, *26* (16), 1735−1745.
(44) Cappellini, E.; Jensen, L. J.; Szklarczyk, D.; Ginolhac, A.; da Fonseca, R. A.; Stafford, T. W.; Holen, S. R.; Collins, M. J.; Orlando, L.; Willerslev, E.; Gilbert, M. T.; Olsen, J. V. *J. Proteome Res.* **2012**, *11* (2), 917−926.
(45) Buckley, M.; Collins, M.; Thomas-Oates, J.; Wilson, J. C. *Rapid Commun. Mass Spectrom.* **2009**, *23* (23), 3843−3854.
(46) Lindahl, T. *Nature* **1993**, *362*, 709−715.
(47) Lindahl, T. *Nature* **1993**, *365*, 700.
(48) Handt, O.; Hoss, M.; Krings, M.; Paabo, S. *Experientia* **1994**, *50* (6), 524−529.
(49) Hoss, M. *Nature* **2000**, *404*, 453−454.
(50) Paabo, S.; Poinar, H. N.; Serre, D.; Jaenicke-Despres, V.; Hebler, J.; Rohland, N.; Kuch, M.; Krause, J.; Vigilant, L.; Hofreiter, M. *Annu. Rev. Genet.* **2004**, *38*, 645−679.
(51) Willerslev, E.; Cooper, A. *Proc. R. Soc. London, Ser. B* **2005**, *272* (1558), 3−16.
(52) Buckley, M.; Walker, A.; Ho, S. Y. W.; Yang, Y.; Smith, C.; Ashton, P.; Oates, J. T.; Cappellini, E.; Koon, H.; Penkman, K.; Elsworth, B.; Ashford, D.; Solazzo, C.; Andrews, P.; Strahler, J.; Shapiro, B.; Ostrom, P.; Gandhi, H.; Miller, W.; Raney, B.; Zylber, M. I.; Gilbert, M. T. P.; Prigodich, R. V.; Ryan, M.; Rijsdijk, K. F.; Janoo, A.; Collins, M. J. *Science* **2008**, *319*, 33c.
(53) 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*.
(54) Sykes, G. A.; Collins, M. J.; Walton, D. I. *Org. Geochem.* **1995**, *23*, 1059−1065.
(55) Collins, M.; Nielsen-Marsh, C.; Hiller, J.; Smith, C.; Roberts, J.; Prigodich, R.; Wess, T.; Csapò, J.; Millard, A.; Turner-Walker, G. *Archaeometry* **2002**, *44* (3), 383−394.
(56) Allentoft, M. E.; Collins, M.; Harker, d.; Haile, J.; Oskam, C. L.; Hale, M. L.; Campos, P. F.; Samaniego, J. A.; Gilbert, M. T. P.; Willerslev, E.; Zhang, G.; Scofield, R. P.; Holdaway, R. N.; Bunce, M. *Proc. R. Soc. London, Ser. B* **2012**, *279* (1748), 4724−4733.
(57) Austin, J. J.; Ross, A. J.; Smith, A. B.; Fortey, R. A.; Thomas, R. H. *Proc. R. Soc. London, Ser. B* **1997**, *264*, 467−74.
(58) Orlando, L.; Ginolhac, A.; Zhang, G.; Froese, D.; Albrechtsen, A.; Stiller, M.; Schubert, M.; Cappellini, E.; Petersen, B.; Moltke, I.; Johnson, P. L. F.; Fumagalli, M.; Vilstrup, J. T.; Raghavan, M.; Korneliussen, T.; Malaspinas, A.-S.; Vogt, J.; Szklarczyk, D.; Kelstrup, C. D.; Vinther, J.; Dolocan, A.; Stenderup, J.; Velazquez, A. M. V.; Cahill, J.; Rasmussen, M.; Wang, X.; Min, J.; Zazula, G. D.; Seguin-Orlando, A.; Mortensen, C.; Magnussen, K.; Thompson, J. F.; Weinstock, J.; Gregersen, K.; Roed, K. H.; Eisenmann, V.; Rubin, C. J.; Miller, D. C.; Antczak, D. F.; Bertelsen, M. F.; Brunak, S.; Al-Rasheid, K. A. S.; Ryder, O.; Andersson, L.; Mundy, J.; Krogh, A.; Gilbert, M. T. P.; Kjaer, K.; Sicheritz-Ponten, T.; Junsen, L. J.; Olsen, J. V.; Hofreiter, M.; Nielsen, R.; Shapiro, B.; Wang, J.; Willerslev, E. *Nature* **2013**, *499* (7456), 74−78.
(59) Bern, M.; Phinney, B. S.; Goldberg, D. *J. Proteome Res.* **2009**, *8*, 4328−4332.

dx.doi.org/10.1021/ac500803w | *Anal. Chem.* XXXX, XXX, XXX−XXX

(60) San Antonio, J. D.; Schweitzer, M. H.; Jensen, S. T.; Kalluri, R.; Buckely, M.; Orgel, J. P. *PLoS One* **2011**, DOI: 10.1371/journal.-pone.0020381.
(61) Huang, S. *Riv. Biol.* **2009**, *102*, 20−22.
(62) Vasan, S.; Zhang, X.; Zhang, X. N.; Kapurniotu, A.; Bernhagen, J.; Teichberg, S.; Basgen, J.; Wagle, D.; Shih, D.; Terlecky, I.; Bucala, R.; Cerami, A.; Egan, J.; Ulrich, P. *Nature* **1996**, *382* (6588), 275−278.
(63) Adam, P.; Philippe, E.; Albrecht, P. *Geochim. Cosmochim. Acta* **1998**, *62* (2), 265−271.
(64) Van Klinken, G. J.; Hedges, R. E. M. *J. Archaeol. Sci.* **1995**, *22* (2), 263−270.
(65) Gupta, N. S.; Michels, R.; Briggs, D. E. G.; Collinson, M. E.; Evershed, R. P.; Pancost, R. D. *Org. Geochem.* **2007**, *38*, 28−36.
(66) McKerrow, J. H.; Robinson, A. B. *Science* **1974**, *183* (4120), 85.
(67) Robinson, A. B.; Robinson, L. R. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88* (20), 8880−8884.
(68) Robinson, N. E.; Robinson, Z. W.; Robinson, B. R.; Robinson, A. L.; Robinson, J. A.; Robinson, M. L.; Robinson, A. B. *J. Pept. Res.* **2004**, *63* (5), 426−436.
(69) 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. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102* (12), 4409−4413.
(70) Humpula, J. F.; Ostrom, P. H.; Ghandi, H.; Strahler, J. H.; Walker, A. K.; Stafford, T. W., Jr.; Smith, J. J.; Voorhies, M. R.; George Corner, R.; Andrews, P. C. *Geochim. Cosmochim. Acta* **2007**, *71*, 5956−5967.
(71) Ostrom, P. H.; Schall, M.; Gandhi, H.; Shen, T. L.; Hauschka, P. V.; Strahler, J. R.; Gage, D. A. *Geochim. Cosmochim. Acta* **2000**, *64* (6), 1043−1050.
(72) Buckley, M. *PLoS One* **2013**, *8*, e59614.
(73) Cleland, T. P.; Voegele, K.; Schweitzer, M. H. *PLoS One* **2012**, *7*, e31443.
(74) Cilia, M.; Fish, T.; Yang, X.; M, M.; Thannhauser, T. W.; Gray, S. *J. Biomol. Technol.* **2009**, *20*, 201−215.
(75) Harper, E.; Kang, A. H. *Biochem. Biophys. Res. Commun.* **1970**, *41*, 482−487.
(76) Seifter, S.; Harper, E. *Methods Enzymol.* **1970**, *19*, 613−635.
(77) Pan, C.; Park, B. H.; McDonald, W. H.; Carey, P. A.; Banfield, J. F.; VerBerkmoes, N. C.; Hettich, R. L.; Samatova, N. F. *BMC Bioinf.* **2010**, *11* (118), 1−14.
(78) Solazzo, C.; Wilson, J.; Dyer, J. M.; Clerens, S.; Plowman, J. E.; von Holstein, I.; Walton Rogers, P.; Peacock, E. E.; Collins, M. J. *Anal. Chem.* **2014**, *86* (1), 567−575.
(79) Li, X.; Cournoyer, J. J.; Lin, C.; O'Connor, P. B. *J. Am. Soc. Mass. Spectrom.* **2008**, *19* (6), 855−864.
(80) Vanacore, R.; Ham, A. J.; Voehler, M.; Sanders, C. R.; Conrads, T. P.; Veenstra, T. D.; Sharpless, K. B.; Dawson, P. E.; Hudson, B. G. *Science* **2009**, *325* (5945), 1230−1234.
(81) Picotti, P.; Aebersold, R. *Nat. Methods* **2012**, *9* (6), 555−566.
(82) Gallien, S.; Duriez, E.; Domon, B. *J. Mass Spectrom.* **2011**, *46* (3), 298−312.
(83) Schweitzer, M. H.; Zheng, W.; Cleland, T. P.; Goodwin, M. B.; Boatman, E.; Theil, E.; Marcus, M. A.; Fakra, S. C., A role for iron and oxygen chemistry in preserving soft tissues, cells and molecules from deep time. *Proc. R. Soc. London, Ser. B* **2014**, *281* (1775).
(84) Medzihradszky, K. F.; Chalkley, R. J. *Mass Spectrom. Rev.* **2013**, 1−21.
(85) Medzihradszky, K. F. *Methods Enzymol.* **2005**, *402*, 209−244.
(86) Bhatia, S. A.; Kil, Y. J.; Ueberheide, B.; Chait, B.; Tayo, L. L.; Cruz, L. J.; Lu, B.; Yates, J. R.; Bern, M. W., Constrained de novo sequencing of peptides with applications to conotoxins. In *RECOMB*; Springer: New York, 2011; pp 16−30.
(87) Avci, R.; Schweitzer, M. H.; Boyd, R.; Wittmeyer, J.; Arce, F.; Calvo, J. *Langmuir* **2005**, *21*, 3584−3591.
(88) Saper, C. B. *J. Histochem. Cytochem.* **2009**, *57*, 1−5.
(89) Bogen, S. A.; Sompuram, S. R., Peptides as immunohistochemistry controls. In *Antigen Retrieval Immunohistochemistry Based Research and Diagnostics*; Shi, S.-R.; Taylor, C. R., Eds.; John Wiley and Sons, Inc: Hoboken, NJ, 2010; pp 123−140.
(90) Witmer, L. M., The extant phylogenetic bracket and the importance of reconstructing soft tissue in fossils. In *Functional Morphology in Vertebrate Paleontology*; Thomason, J. J., Ed.; Cambridge University Press: Cambridge, 1995.
(91) Gilbert, M. T. P.; Bandelt, H.-J.; Hofreiter, M.; Barnes, I. *Trends in Ecology and Evolution* **2005**, *20* (10), 541−544.
(92) Collins, M. J.; Riley, M.; Child, A. M.; Turner-Walker, G. *J. Archaeol. Sci.* **1995**, *22*, 175−183.
(93) Collins, M. J.; Waite, E. R.; van Duin, A. C. T. *Philos. Trans. R. Soc., B* **1999**, *1999* (354), 51−64.
(94) Collins, M. J.; Gernaey, A.; Nielsen-Marsh, C. M.; Vermeer, C.; Westbroek, P. *Geology* **2000**, *28* (12), 1139−1142.
(95) Hofreiter, M.; Serre, D.; Poinar, H. N.; Kuch, M.; Paabo, S. *Nat. Rev. Genet.* **2001**, *2* (5), 353−359.
(96) Trueman, C. N.; Palmer, M. R.; Field, J.; Privat, K.; Ludgate, N.; Chavagnac, V.; Eberth, D. A.; Cifelli, R.; Rogers, R. R. *C. R. Palevol* **2008**, *7*, 145−158.
(97) Schweitzer, M. H.; Wittmeyer, J. L.; Horner, J. H.; Toporski, J. B. *Science* **2005**, *307*, 1952−1955.
(98) Weiner, S.; Wagner, H. D. *Annu. Rev. Mater. Sci.* **1998**, *28*, 271−298.
(99) Schweitzer, M. H.; Wittmeyer, J. L.; Horner, J. R. *Proc. R. Soc. London, Ser. B* **2007**, *274*, 183−197.
(100) Pevzner, P. A.; Kim, S.; Ng, J. *Science* **2008**, *321*, 104b.
(101) Asara, J. M.; Schweitzer, M. H. *Science* **2008**, *319*, 33d.
(102) Briggs, D. E. G. *Eclogae Geol. Helv.* **1995**, *88* (3), 623−626.
(103) Briggs, D. E. G.; Evershed, R. P.; Lockheart, M. J. *Paleobiology* **2000**, *26* (4), 169−193.
(104) Schweitzer, M. H.; Watt, J. A.; Avci, R.; Forster, C. A.; Krause, D. W.; Knapp, L.; Rogers, R. R.; Beech, I.; Marshall, M. *J. Vertebr. Paleontol.* **1999**, *19* (4), 712−722.
(105) Schweitzer, M. H.; Watt, J. A.; Avci, R.; Knapp, L.; Chiappe, L.; Norell, M.; Marshall, M. *J. Exp. Zool.* **1999**, *285*, 146−157.
(106) Schweitzer, M. H.; Marshall, C. L. *J. Exp. Zool.* **2001**, *291*, 317−338.
(107) Honda, K.; Casadesus, G.; Petersen, R. B.; Perry, G.; Smith, M. A. *Ann. N.Y. Acad. Sci.* **2004**, *1012*, 179−182.
(108) Rajotte, D.; Arap, W.; Hagedorn, M.; Koivunen, E.; Pasqualini, R.; Ruoslahti, E. *J. Clin. Invest.* **1998**, *102* (2), 430−437.
(109) Gerber, H.-P.; Vu, T. H.; Ryan, A. M.; Kowalski, J.; Werb, Z.; Ferrara, N. *Nat. Med.* **1999**, *5* (6), 623−628.
(110) Chi, B.; Snow, J. Z. *Project Exploration 10-Year Retrospective Program Evaluation: Summative Report*; University of California: Berkeley, 2010; pp 1−82.
(111) Van der Plas, A.; Nijweide, P. J., Isolation and purification of osteocytes. *J. Bone Miner. Res.* **1992**, *7* (389−396).
(112) Westbroek, I.; deRooij, K. E.; Nijweide, P. J. *J. Bone Miner. Res.* **2002**, *17* (5), 845−853.