Phylogenetic and physiological signals in metazoan fossil biomolecules

Jasmina Wiemann1*, Jason M. Crawford2,3,4, Derek E. G. Briggs1,5

Proteins, lipids, and sugars establish animal form and function. However, the preservation of biological signals in fossil organic matter is poorly understood. Here, we used high-resolution in situ Raman microspectroscopy to analyze the molecular compositions of 113 Phanerozoic metazoan fossils and sediments. Proteins, lipids, and sugars converge in composition during fossilization through lipoxidation and glycoxidation to form endogenous N-, O-, and S-heterocyclic polymers. Nonetheless, multivariate spectral analysis reveals molecular heterogeneities: The relative abundance of glycoxidation and lipoxidation products distinguishes different tissue types. Preserved chelating ligands are diagnostic of different modes of biominalization. Amino acid–specific fossilization products retain phylogenetic information and capture higher-rank metazoan relationships. Molecular signals survive in deep time and provide a powerful tool for reconstructing the evolutionary history of animals.

INTRODUCTION
Proteins, lipids, and sugars are the major building blocks of animal tissues. These structural biomolecules store a wealth of biological information at different levels of hierarchical organization (1, 2). On the tissue scale, omnipresent proteins co-occur with lipids and sugars in amounts diagnostic of taxonomy and anatomical feature (1, 2). Proteins represent direct translation products of the genetic code and are therefore compositionally more diverse than core metabolites, such as sugars and lipids (3). This diversity adds biological information at a supramolecular level: Interactions between coordinating amino acid residues exposed on the protein surface and biominerals determine the properties of a biomaterial (4). On a molecular level, phylogenetic information is stored in the amino acid sequence of a protein (3). When analyzed in a comparative framework, structural biomolecules reveal information about physiology, tissue types, and the nature of unique biocomposite materials, and they resolve relationships among animals. However, these inferences have been largely limited to extant or recently extinct taxa (5).

Until recently, biomolecules were thought to decay rapidly post-mortem (6–8), and research on physiology, biomineralization, and relationships in extinct metazoans relies largely on the morphology of fossils, which is often incomplete, distorted, or, for some taxa, completely absent. Proteins, lipids, and sugars can fossilize through oxidative cross-linking (9). This reaction scheme transforms and stabilizes structural biomolecules under chemically oxidative conditions during early diagenesis (9) and differs in its reaction educts and products from those of sulfurization processes that primarily alter lipids under highly reducing sediment conditions (10). Proteins, lipids, and sugars are transformed into advanced glycoxidation and lipoxidation end products (AGEs and ALEs) composed of N-, O-, and S-rich heterocyclic polymers, which are water insoluble, largely inaccessible to microbial decay, and chemically stable (9, 11).

This fossilization pathway is centered around suitable nucleophilic amino acid residues in a protein reacting with reactive carbonyl species (RCS) formed during early diagenesis (9). These RCS can be derived from either lipids or sugars (12, 13), thus connecting protein information to core metabolites. Lipoxidation and glycoxidation processes are generally known to occur under alkaline pH conditions and are catalyzed by abundant transition metal ions and dissolved phosphate in the pore water (9, 12, 14). However, this model of biomolecule fossilization has only been tested in vertebrate hard tissues (9, 13, 14), and neither patterns in the composition of N-, O-, or S-heterocyclic polymers nor their potential as biomarkers has been explored. Obtaining even partial information on molecular biological signals from fossils would foster their integration into molecular datasets and thereby fundamentally advance our understanding of the evolution of animal life.

MATERIALS AND METHODS
High-resolution in situ Raman microspectroscopy setup and protocol
We used high-resolution in situ Raman microspectroscopy, a nondestructive method (15, 16), to analyze biomolecule fossilization products in 96 specimens of extant (n = 20) and representative fossil (n = 76) metazoans and their associated sediments (in 17 cases) (Figs. 1 to 3). From this dataset (section S1 and Supplementary Data), we identified patterns in composition and biological importance by statistically analyzing different aspects of the spectrally encoded information on biomolecules and their fossilization products. The analyzed fossil samples [selected on the basis of (9)] range in age from Cambrian to Tertiary and provide a taxonomic coverage of major animal groups with a notable fossil record (Supplementary Data).

All samples were individually surface cleaned with 70% ethanol and analyzed (section S2) using a Horiba LabRAM HR800 (532 nm, 20 mW, 1800 grooves/mm grating, 10 s, 10 technical replicates, 500 to 2000 cm⁻¹). The spectra were obtained and processed in LabSpec 5. All spectra were then baselined, smoothed, and analyzed in SpectraGryph 1.2. Individual Raman bands were identified from the spectra through an automated peak search (yielding Supplementary Data, band sets 1 to 4).

1Department of Earth and Planetary Sciences, Yale University, New Haven, CT 06511, USA. 2Department of Chemistry, Yale University, New Haven, CT 06520, USA. 3Department of Microbial Pathogenesis, Yale University, New Haven, CT 06536, USA. 4Chemical Biology Institute, Yale University, West Haven, CT 06516, USA. 5Yale Peabody Museum of Natural History, Yale University, New Haven, CT 06511, USA.
*Corresponding author. Email: jasmina.wiemann@yale.edu
Three-dimensional ChemoSpace plot and discriminant analysis: Organic diagenesis

We selected 35 Raman bands (17) (sections S2 and S3 and Supplementary Data, band set 1) from the organic fingerprint region (500 to 1800 cm\(^{-1}\) Raman shift) to characterize the change in total organic composition (\(n = 96\) modern and fossil specimens; Supplementary Data) evident in averaged spectra for eggshells, bones, teeth/enamel scales, vertebrate soft tissues, biomineralized invertebrate tissues, and invertebrate soft tissues during fossilization. Relative spectral intensities were averaged for tissue categories in SpectraGryph 1.2, transformed into a variance-covariance matrix, and subjected to a ChemoSpace principal components analysis (PCA) in PAST 3. The data for each averaged tissue category in the resulting two-dimensional (2D) ChemoSpace (\(x\) and \(y\) coordinates) were complemented with the number of significant peaks per spectrum (\(z\) coordinate) representing distinct organic functional groups (automated peak search; section S2). On the basis of these Cartesian coordinates, a 3D plot (Fig. 1A) illustrating the diversity in organic functions versus the distinctness of the composition of each sample (fossil or modern) was generated. This dataset, with the addition of a single parameter assigning a modern or fossil identity, was subjected to a discriminant analysis in PAST 3 (Fig. 1A).

**Fig. 1.** Changes in the molecular composition of animal tissues through geologic time. (A) ChemoSpace PCA and discriminant analysis (DA) of in situ Raman spectra of modern and fossil animal samples (\(n = 96\)) averaged by individual tissue types (band set 1). PC 1 and PC 2 define a principal component space on which a measure of functional diversity (\(z\) axis) is plotted for each modern and fossil animal tissue type (see Materials and Methods). Blue columns represent unaltered biomolecules in modern tissues (corresponding to the blue convex hull), and orange columns represent biomolecular fossilization products in fossil tissues (orange convex hull). Modern tissues remodeled in vivo are labeled. The functional diversity of structural biomolecules in modern tissues is high, whereas that in different fossil tissues is much lower. The red vector arrow indicates the general trajectory (recovered in PCA (Principal Component Analysis) and DA) in molecular composition followed by samples during fossilization and identifies the processes of glycoxidation and lipoxidation. (B to D) Structural biomolecules in modern tissues with labeled functional groups. (B) Protein, a tripeptide with the amino acids lysine, cysteine, and valine. (C) Lipid. (D) Sugar (the glycosaminoglycan chitin). The colored circles indicate extensions of the biopolymer.
Fig. 2. High-resolution in situ Raman spectra of fossil animal tissues and assessment of their diagenetic alteration. (A) Plot of Raman spectra of fossil metazoans in light blue; averaged spectra: black, all fossil metazoans (n = 76); dotted blue, vertebrates (n = 41); and dotted red, invertebrates (n = 35). Red arrows: advanced lipoxidation/glycoxidation end products ALEs/AGEs; R, variable residues. Units evident in fossil tissues are displayed at the top. (B) Bar charts, averaged for tissue type (Supplementary Data), show normalized signal intensities at 1685 cm⁻¹: blue, trans-amide, peptide bond (A_p); and at 1701 cm⁻¹: yellow, cis-amide, diagenetically formed (A_d). Ratios (A_p/A_d) are based on these signals. Asterisks, pristine preservation of the protein/PFP phase: A_p/A_d > 1. Yellow/blue circles, extension of the polymer. (C) Molecular mechanisms for the transformation of structural biomolecules during fossilization. Cat, catalyst; Nu, nucleophile; pRCS, primary RCS; Δ, thermal energy input. Blue circles, extension of the polymer. (D) Discriminant analysis of organic matter in fossil animal tissues (n = 76; green dots) and sediments (n = 17; gray dots) based on Raman intensities at 24 band positions (Supplementary Data). Green/blue arrows, trajectory of fossil/sediment samples. Discriminant factors are listed. a.u., arbitrary units.
trans-A and cis-A by its peptide bonds [cross-linking involves the peptide skeleton, which is characterized to affect only suitable amino acid residues (Fig. 2C), advanced lipid- or sugar-derived RCS.

A novel proxy to quantify diagenetic alteration: The trans/cis-amide ratio

The degree of protein transformation in fossil metazoa provides an estimate of diagenetic alteration (Fig. 2). Proteins cross-link with lipid- or sugar-derived RCS. Whereas initial cross-linking reactions affect only suitable amino acid residues (Fig. 2C), advanced cross-linking involves the peptide skeleton, which is characterized by its peptide bonds [trans-amides (A\textsubscript{T}); Fig. 2B] (12). Advanced cross-linking tends to produce heterocycles through condensation reactions, which yield characteristic cis-amide (A\textsubscript{C}) signals in the Raman spectra of fossil animal samples. Cis-amides in fresh proteins, lipids, and sugars yield a very weak Raman signal. The ratio of intact peptide bonds (trans-amides) to diagenetic reaction products (cis-amides) provides an assessment of the degree of diagenetic alteration of structural biomolecules in fossil animal samples (Fig. 2B). Relative signal intensities for the bands at 1685 cm\textsuperscript{-1}, representing trans-amides, and 1702 cm\textsuperscript{-1}, representing cis-amides, were extracted from averaged spectra for the different fossil metazoan tissue types (n = 76), and the ratio of trans/cis-amides (Supplementary Data, specimen data) was plotted (Fig. 2A, bar charts). Values of (A\textsubscript{T}/A\textsubscript{C}) > 1 indicate that more peptide bonds are intact than are diagnostically altered.

Discriminant analysis: Fossil versus sediment organic matter

The possibility of contamination with exogenous organic matter is a common concern where signals indicate intact peptide bonds in fossils (Fig. 2, A and B). We test this by determining signal intensities at 24 band positions characterizing fossil organic matter (Fig. 2A) in the spectra of all metazoan fossils (n = 76) and associated sediment samples (n = 17). The selected bands correspond to a range of different organic functional groups (Supplementary Data, band set 2). Relative intensities in SpectraGryph 1.2 were transformed into a variance-covariance matrix, identified as fossil or sediment, and subjected to a discriminant analysis in PAST 3 (Fig. 2D). Discriminant factors between fossil and sediment were identified (Supplementary Data, band set 2).

Two-dimensional ChemoSpace PCA and discriminant analysis: Tissue-type clustering

A discriminant analysis of fossil samples was carried out to explore the possibility and nature of a preserved tissue-specific signature (Figs. 1A and 2A). The spectra (n = 76 taxa) were converted into a variance-covariance matrix and identified as biomineralized or non-biomineralized. The resulting data matrix was subjected to a discriminant analysis in PAST 3 (Fig. 3A). In addition, the 24 Raman band intensities selected for fossil metazoa (n = 76) were subjected to a ChemoSpace PCA in PAST 3 without the parameter discriminating biomineralized and non-biomineralized samples (Fig. 3A).

A second analysis selected vertebrate hard tissues (n = 41) from the matrix used for the analyses shown in Fig. 3A (n = 76). Samples were identified as eggshells, bones, and teeth excluding conodonts (Fig. 2A). The data matrix incorporating these additional parameters was subjected to a discriminant analysis in PAST 3. The 24 Raman band intensities selected for fossil metazoa (n = 76) were subjected to a ChemoSpace PCA in PAST 3 without distinguishing between eggshell, bone, and teeth. Principal component (PC) 1 (43%) and PC 2 (31%) separate samples characterized by features identified by the corresponding discriminant analysis (Fig. 3B). Convex hulls show no overlap in the molecular composition of calcite- and apatite-biomineralized fossil samples (Fig. 3B).

Spectral cluster analysis: Preserved phylogenetic signals in different tissue types

Phylogenetic information is stored in the amino acid composition of a protein, and different tissues contain different key structural proteins. Therefore, samples were grouped by tissue types. Because compositional heterogeneities resulting from different reacting amino acids are expected to be minimal, potential heterogeneities resulting from different modes of preservation need to be eliminated; therefore, relative spectral intensities at 550 cm\textsuperscript{-1} (S-heterocycles) and 1580 cm\textsuperscript{-1} (N-heterocycles) Raman shifts were determined for samples of every tissue category in the data matrix generated for the
ChemoSpace analysis (Fig. 3A). The ratio of S- and N-heterocycles within a tissue category reflects the original ratio of thiol- to amine-bearing amino acid residues (fig. S1), and outliers in the tissue-specific $[\text{C-S}] / [\text{C-N}]$ ratios were omitted. Within each tissue category, samples of identical or directly comparable lithologies were selected to avoid signal distortion through minute differences in the mode of organic matter preservation. Taphonomically comparable fossils (Supplementary Data, specimen data)—eggshells ($n = 8$), teeth ($n = 6$), bones ($n = 10$), biomineralized invertebrate samples ($n = 7$), and non-biomineralized invertebrate samples ($n = 6$)—were used to determine signal intensities at 20 band positions (Supplementary Data, band set 4) characterizing the specific composition of amino acid–derived N- and S-heterocycles (fig. S1) in fossil organic matter.

Relative intensities in SpectraGryph 1.2 were transformed into a variance-covariance matrix, which was the basis for two different discriminant analyses of biomineralized ($n = 56$, blue dots) and non-biomineralized ($n = 20$, yellow dots) samples, based on the same Raman data. Black outlined dots, vertebrate samples (generally more ALEs, compare Table 1); red outlined, invertebrate samples (generally fewer ALEs than AGEs, compare Table 1); $\Delta$, thermally matured samples (Cambrian Burgess Shale); blue arrow, trajectory of biomineralized tissues; yellow arrow, that of no-biomineralized tissues. (B) Tissue-type signal. Data on vertebrate eggshells ($n = 14$), bones ($n = 19$), and teeth excluding conodonts ($n = 8$) from the data set used in (A) were subjected to a discriminant analysis. Green, eggshells; dark blue, bones; and light blue, teeth.

Fig. 3. Clustering of fossil animal tissues. (A) Biomineralization signal for all metazoan fossils based on spectral intensities at 24 Raman band positions ($n = 76$) (see Materials and Methods) complemented by a discriminant analysis of the biomineralized ($n = 56$, blue dots) and non-biomineralized ($n = 20$, yellow dots) samples, based on the same Raman data. Black outlined dots, vertebrate samples (generally more ALEs, compare Table 1); red outlined, invertebrate samples (generally fewer ALEs than AGEs, compare Table 1); $\Delta$, thermally matured samples (Cambrian Burgess Shale); blue arrow, trajectory of biomineralized tissues; yellow arrow, that of no-biomineralized tissues. (B) Tissue-type signal. Data on vertebrate eggshells ($n = 14$), bones ($n = 19$), and teeth excluding conodonts ($n = 8$) from the data set used in (A) were subjected to a discriminant analysis. Green, eggshells; dark blue, bones; and light blue, teeth.
The composition and endogeneity of metazoan biomolecule fossilization products

The composition of the fossil organic phase was characterized by plotting all fossil spectra (Fig. 2A) and superimposing the average for fossil metazoans and for the subset of vertebrates and invertebrates. All fossil spectra share key features, which correspond to N-, O-, and S-rich heterocyclic polymers. The Raman bands identify signals associated with thioethers, ethers, esters, carboxyls, and cis- and trans-amides, as well as a suite of heterocycles [compare structures in Figs. 1 (B to D) and 2 (A and B) and in fig. S1].

Sulfurization does not generally occur under oxidative sediment conditions (10) but was considered as a potential source of thioethers and S-heterocycles. Tests of covariation of the organo-sulfur abundance in fossil and corresponding extant tissues (section S5 and figs. S2 and S3) and multivariate assessments of the organo-sulfur species distribution in fossil and sediment organic matter indicate that early diagenetic sulfurization (fig. S3, A and B) does not account for the thioethers and S-heterocycles detected in our samples. A plot (fig. S1) of the net compositional change during fossilization identifies oxidative cross-linking, specifically glycoxidation and lipoxidation, as the early diagenetic process (section S4). Oxidative cross-linking is shown to have occurred even in pressure- and temperature-matured Burgess Shale fossils (n = 4 specimens), although the intensity of spectral signatures for N-, O-, and S-heterocycles is reduced in these samples, while aromatic compounds are relatively more abundant.

The presence of trans-amides (Fig. 2B) in fossils supports the preservation of peptide bonds, diagnostic of proteins, whereas cis-amides are common in AGEs and ALEs (12). The lack of cis-amide Raman signals in fresh animal tissues indicates that the fossil signals are diagenetic in origin (Figs. 1A and 2B and fig. S1). Thus, the ratio of trans- to cis-amides (Fig. 2B) was calculated as a measure of diagenetic transformation (section S2 and Supplementary Data). Trans-amides are more abundant than cis-amides in fossil eggshell, vertebrate soft tissues, and invertebrate biomineralized tissues (Fig. 2B).

The preservation of peptide bonds could be a result of exceptional preservation or exogenous contamination. We tested for contamination by comparing signal intensities of the spectra of all fossil metazoan samples with associated sediment samples. Organic matter in fossils (Fig. 2D) contains abundant signals for trans-amides (i.e., peptide bonds), thioethers, and S-heterocycles, whereas that in sediment is characterized by substantially more peroxidized aromatics and simple aliphatic compounds. The distinct differences in the composition of fossil animal and sediment organic matter exclude contamination.

Although metazoan structural biomolecules converge toward N-, O-, and S-heterocyclic polymers as a result of diagenetic transformation (Figs. 1 and 2), the average molecular composition of fossil vertebrate and invertebrate tissues is distinctly different (Fig. 2A), which shows that tissue-specific signals survive fossilization (section S6). Even the trans/cis-amide ratios (Fig. 2B), which reflect the degree of diagenetic alteration, are specific to particular tissue categories.

Tissue-type signals

The key differences between different tissue categories were determined with additional discriminant and ChemoSpace PCAs (Fig. 3). The resulting variance-covariance matrices were analyzed
for features discriminating between biomineralized and non-biomineralized samples (Fig. 3A) and between eggshell, teeth (excluding conodonts), and bone samples (Fig. 3B). There is minimal overlap between the convex hulls of biomineralizing and non-biomineralizing metazoans in the ChemoSpace PCA (Fig. 3A). PC eigenvectors correspond to key differences identified by the discriminant analysis. PC 1 separates fossil biomineralizers and non-biomineralizing animals. Organic matter in fossil biomineralizers is characterized by abundant peptide bonds (trans-amides) and coordinating ligands such as N-, O-, and S-heterocycles. PC 1 also incorporates ALE abundance (evidenced by increased amounts of thioethers): Invertebrates generally yield negative PC 1 values, as glycoxidation (based on chitin) signals appear to be favored over lipoxidation (section S5). Thermally matured samples such as those from the
Cambrian Burgess Shale (Fig. 3A) still group with non-biomineralized invertebrates, suggesting that tissue-type signals survive the effects of metamorphism to a certain extent. The ChemoSpace of eggshell, teeth, and bone samples (Fig. 3B) yielded two distinct groups: eggshells (calcite) and a largely overlapping group of bones and teeth (apatite). The organic phase in fossil eggshell calcite is distinguished by more abundant peptide bonds and S- and O-heterocycles, whereas the organic phase of fossil apatite is characterized by N-heterocycles, which are more abundant in bones than in teeth. Calcite and apatite biomarkers also differ in the context of coordinating ligands in the organic phase, suggesting that amino acid–biomineral interactions specific to different hard tissues survive fossilization (Fig. 3B). The preservation of peptide bonds (trans-amides) within different biomarkers (compare Figs. 2B and 3B) reflects the degree to which organic matrix proteins were shielded from diagenetic peroxidation.

Phylogenetic and metabolic signals in metazoan biomolecule fossilization products

The subset of fossil metazoans suitable for phylogenetic clustering (see Materials and Methods and sections S2 and S6) includes eggshell, bone, teeth (excluding conodonts), biomineralized invertebrate, and non-biomineralized invertebrate samples. Outliers in the tissue-specific [(C=S)/(C=N)] ratios are listed in section S7 and Supplementary Data. In addition, a hierarchical cluster analysis was performed, including all tissue types, except eggshells, the siliceous sponge, and Lithostrotion. These RCS can promote oxidative cross-linking of cysteinyl residues (amino acid) with lipid-derived RCS (Figs. 1A and 2C and fig. S1) (9, 12, 26). The different tissue categories of metazoan fossils are distinguished by original differences in the amounts of proteins, lipids, and sugars (Fig. 1), which transform into corresponding AGEs and ALEs during fossilization. Certain functional groups are lost (Fig. 1A), while the formation of thiocysteines, N-, O-, and S-heterocycles, carbonyls, ethers, and esters is favored (Figs. 1A and 2A). Diagenesis results in convergence toward chemically stable, cross-linked units, which have the potential to survive in deep time (Figs. 1A and 2).

N-, O-, and S-heterocyclic polymers and aliphatic and aromatic compounds are detected even in fossils from the Burgess Shale (Fig. 2A), which have undergone temperature and pressure maturation. These carbonaceous remains are associated with clay minerals, which generate RCS on their surfaces (28). These RCS can promote the oxidative conditions required for early diagenetic glycoxidation and lipoxidation of adjacent biomolecules, even where environmental conditions do not promote oxidative chemistry.

Dense, avascular eggshell calcite offers most protection against alteration of organic components (Figs. 2B, 3B, and 4) in contrast to vascularized, phosphatic biomineralized bone, teeth, and enamel scales (Figs. 2B, 3B, and 4); compare Fig. 1A). This contrast is reflected in the quality of the phylogenetic signal in animal fossil organic matter (Fig. 4). However, differences in the composition of fossil organic matter across tissues reflect not only differential protection by minerals (Figs. 3, A and B, and 4) but also the nature of oxidative cross-links where the main source of RCS may be lipid or sugar, depending on the original tissue composition (Fig. 3, A and B, and Table 1). Fossil organic matter preserves coordinating ligands diagnostic of the biomineral phase (Fig. 3, A and B) even where hard parts have been lost, altered, or replaced through pore water interactions.

Our data [Figs. 1A, 2 (A and B), and 4 (A to D) and fig. S1] suggest that amino acids cross-link with RCS in particular ways during early diagenesis, “fixing” the protein composition in amino acid–specific fossilization products (Fig. 5). Organic matter in fossil eggshells, and in biomineralized and non-biomineralized invertebrate tissues, preserves the most reliable phylogenetic signals and correctly resolves higher-rank metazoan relationships. Although the retention of chemotaxonomic signals in lipid biomarkers associated with fossils has previously been noted (29–32), the application of phylogenetic signals based on amino acid transformations in fossil (insoluble) organic matter represents an entirely new approach.

Metazoan biomolecule fossilization products diagnostic of glycoxidation and lipoxidation show that early diagenetic cross-linking of structural biomolecules (12, 26, 33) not only occurs in vertebrate
tools to address fundamental questions on the evolution of bio-
framework. These comparative molecular analyses provide powerful
and phylogenetic affinities when considered in a comparative
Biomolecule fossilization products not only reveal the diagenetic
preserves biological signals regardless of age or diagenetic alteration.

CONCLUSIONS
Fossil organic matter formed through oxidative cross-linking
preserves biological signals regardless of age or diagenetic alteration.
Biomolecule fossilization products not only reveal the diagenetic
history of an animal fossil but also illuminate its structural nature
and phylogenetic affinities when considered in a comparative
framework. These comparative molecular analyses provide powerful
tools to address fundamental questions on the evolution of bio-
mineralized tissues and have the potential to resolve the position of fossils in the animal tree of life.

SUPPLEMENTARY MATERIALS
Supplementary material for this article is available at http://advances.sciencemag.org/cgi-
content/full/6/28/eaba6883/DC1

Identification of amino acids

REFERENCES AND NOTES
4. L. Addadi, S. Weiner, Interactions between acidic proteins and crystals: Stereochemical
5. D. E. G. Briggs, R. E. Summons, Ancient biomolecules: Their origins, fossilization, and role
B. Shapiro, P. Ostrom, H. Gandhi, W. Miller, B. Raney, M. I. Zyliber, M. T. P. Gilbert,
R. V. Prigodich, M. Ryan, K. F. Rjidsj, A. Janoo, M. J. Collins, Weighing the mass spec-
metric evidence for authentic Tyrannosaurus rex collagen. Science 319, 33
(2008).
J. Proteome Res. 8, 4328–4332 (2009).
8. B. Demarchi, S. Hall, T. Roncal-Herrero, C. L. Freeman, J. Woolley, M. K. Crisp, J. Wilson,
A. Fotakis, R. Fischer, B. M. Kessler, R. R. Jerse-Christensen, J. V. Olsen, J. Halte, J. Thomas,
C. W. Marean, J. Parkington, S. Presslee, J. Lee-Thorp, P. Ditchfield, J. F. Hamilton,
M. W. Ward, C. M. Wang, M. D. Shaw, T. Harrison, M. Domínguez-Rodrigo, R. D. E. MacPhee,
A. Kwokson, M. Ecker, L. K. Horwitz, M. Chazan, R. Kröger, J. Thomas-Oates, J. H. Harding,
E. Cappellini, K. Penkman, M. J. Collins, Protein sequences bound to mineral surfaces
persist into deep time. eLife 5, e17092 (2016).
Fossilization transforms vertebrate hard tissue proteins into N-heterocyclic polymers.
10. P. Adam, P. Schneckenburger, P. Schaeffer, P. Albrecht, Clues to early diagenetic
sulfurization processes from mild chemical cleavage of labile sulfur-rich
Volatile compounds in archaeological plant remains and the Maillard reaction during
12. G. Vitolli, D. De Maddis, A. Cipak, N. Zarkovic, M. Carini, G. Aldini, Advanced glycoxidation
and lipoxidation end products (AGEs and ALEs): An overview of their mechanisms


27. J. Wiemann, M.A. Theurer, Determining the distinction between organic compounds found in fossil plants versus fossil animals via Raman spectroscopy, in GSA Annual Meeting in Phoenix, Arizona (Geological Society of America, 2019).


Acknowledgments: We are grateful to J. Gauthier, M. Norell, and P. Hull for helpful discussions and sample materials. P. Mayer and B. Simpson provided specimens. S. Butts and J. Ustrup helped with the selection of suitable invertebrates from the collections of the Yale Peabody Museum. M. Fabbri provided helpful comments. V. Rhäe catalogued all vertebrate specimens. Funding: This research was supported by graduate student research grants from the Yale Institute for Biospheric Studies, the Geological Society of America, and by the Yale Peabody Museum Invertebrate Paleontology Division. Author contributions: J.W., J.M.C., and D.E.G.B. designed the research and discussed analytical procedures. J.W. selected sample materials, analyzed samples using Raman spectroscopy, designed and performed the statistical analyses, and prepared the figures. All authors discussed the results, wrote, and reviewed the manuscript. Competing interests: The authors declare that they have no competing interests. Data and materials availability: All data supporting the findings of this study are available within the main article (Figs. 1 to 4 and Table 1), the Supplementary Materials, and the Source Data (Figs. 3, A and B, and 4A). The Supplementary Materials include specimen details, method details, Raman band assignments (for band sets 1 to 4), details on the band selection, data on the observed chemical transformations during fossilization (additional to Fig. 1A), tests for input from sulfurization processes, and information on the subsampling and outlier analysis performed for the phylogenetic clustering. Source data used for the 3D ChemoSpace (Fig. 1A), the biominalization ChemoSpace analysis (including PC loadings; Fig. 3A), the tissue-type ChemoSpace analysis (including PC loadings; Fig. 3B), and the phylogenetic clustering of different tissues (Fig. 4A) are available in the Supplementary Data spreadsheet and in the data files. Sample materials can be made available from the corresponding author upon request.

Submitted 23 December 2019
Accepted 1 June 2020
Published 10 July 2020
10.1126/sciadv.aba6883

Phylogenetic and physiological signals in metazoan fossil biomolecules
Jasmina Wiemann, Jason M. Crawford and Derek E. G. Briggs

Sci Adv 6 (28), eaba6883.
DOI: 10.1126/sciadv.eaba6883