PALEONTOLOGY

Phylogenetic and physiological signals in metazoan fossil biomolecules

Jasmina Wiemann[1]*, Jason M. Crawford[2,3,4], Derek E. G. Briggs[1,5]

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 biomineralization. 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). 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 postmortem (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 SpectraGraph 1.2. Individual Raman bands were identified from the spectra through an automated peak search (yielding Supplementary Data, band sets 1 to 4).

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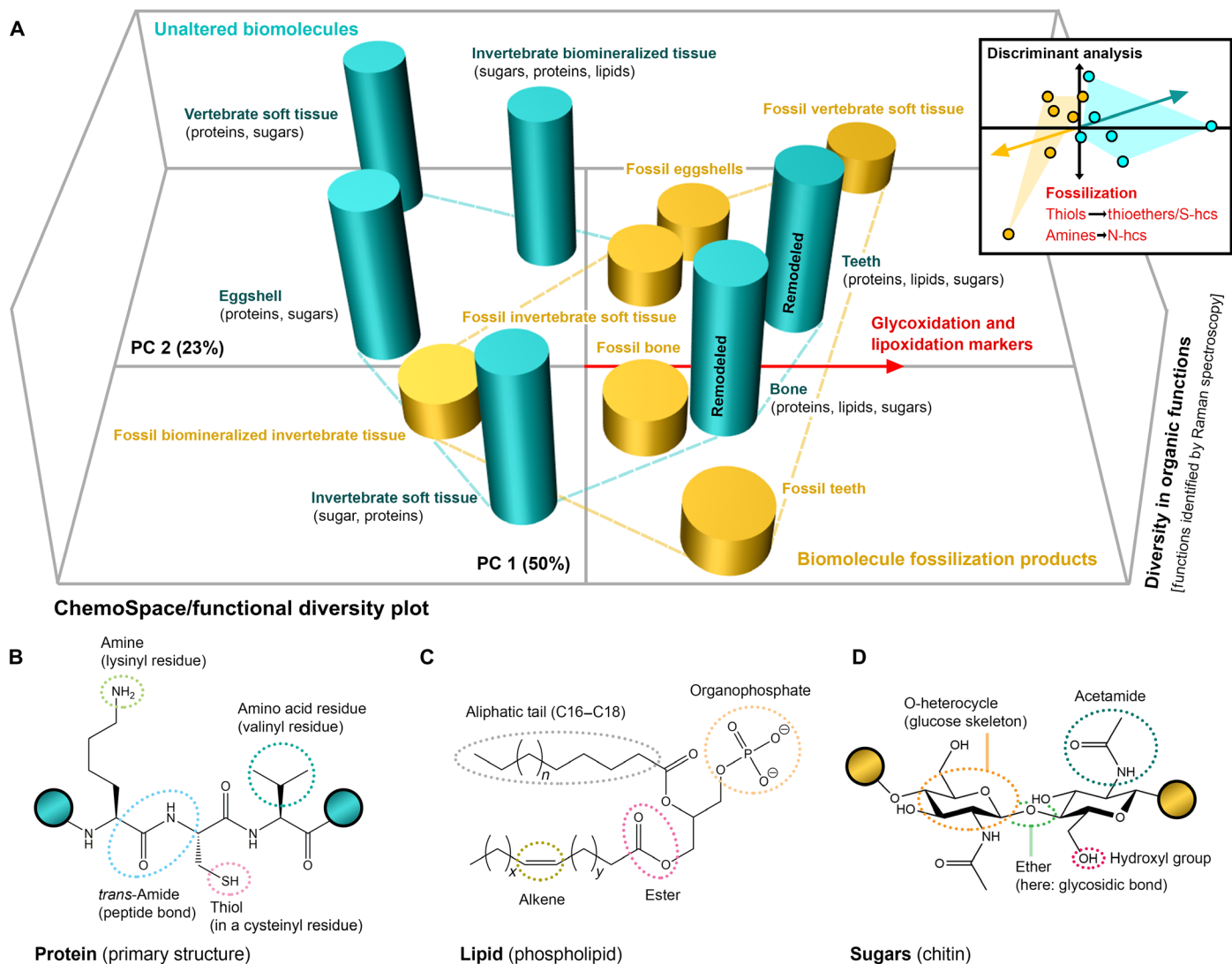*Corresponding author. Email: jasmina.wiemann@yale.edu

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.

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).
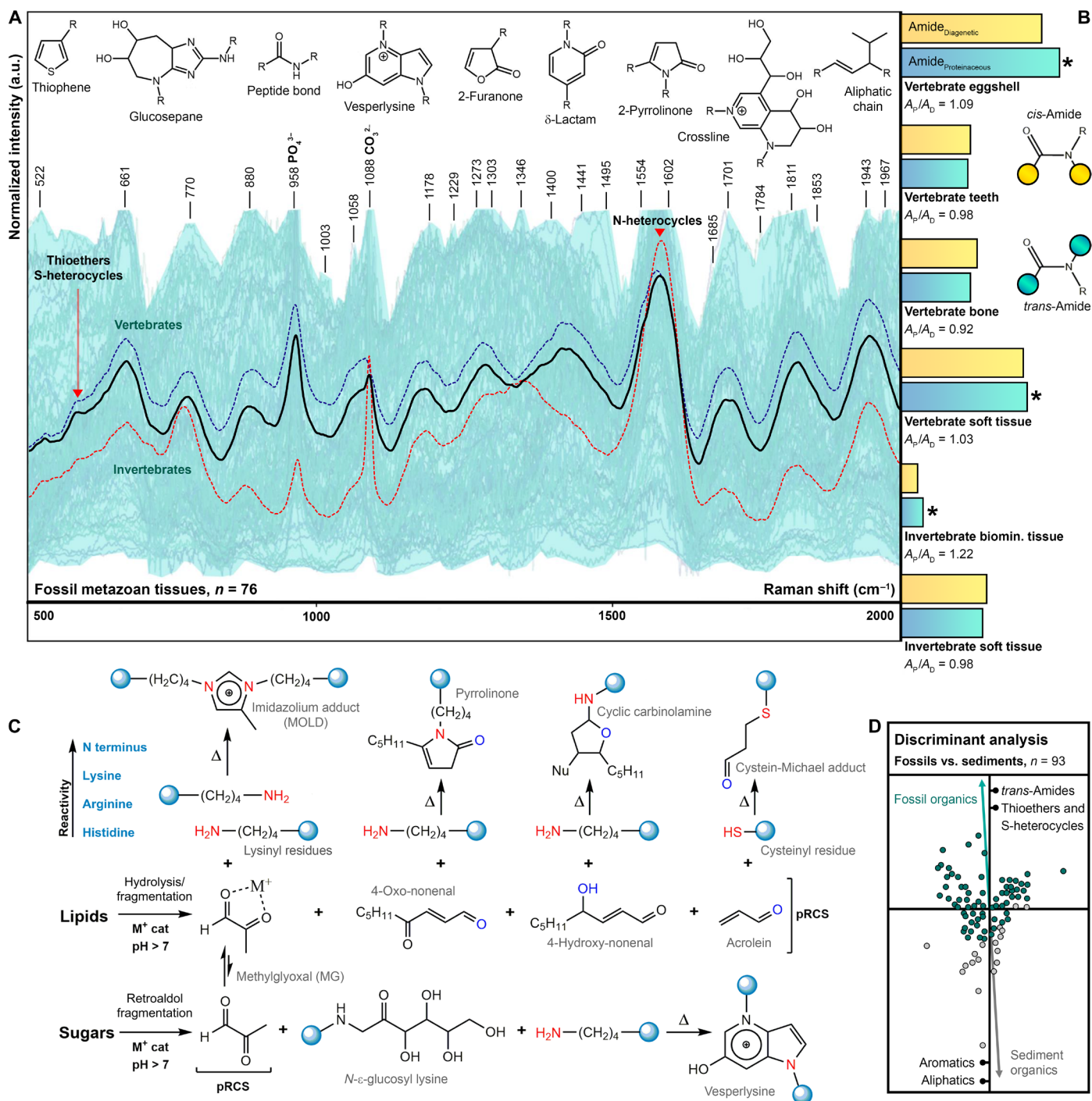
Fig. 2. High-resolution in situ Raman spectra of fossil animal tissues and assessment of their diagenetic alteration. (**A**) Plot of Raman spectra of fossil metazoans in light blue; averaged spectra: black, all fossil metazoans ($n = 76$); dotted blue, vertebrates ($n = 41$); and dotted red, invertebrates ($n = 35$). 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/gray arrows, trajectory of fossil/sediment samples. Discriminant factors are listed. a.u., arbitrary units.

Table 1. Summary data related to fossilization. The six tissue types considered here are compared in terms of composition (the relative in vivo abundance of structural biomolecules listed in decreasing order). Avascular tissues that are not remodeled in vivo retain a high phylogenetic signal. Tissues retaining a weak phylogenetic signal, in contrast, suffer from in vivo tissue remodeling. The ratio of S- to N-heterocycles ([—C—S—]/[—C—N—]) in fossils is quantified for eggshells ($n = 14$), teeth including conodonts ($n = 14$), bones ($n = 19$), vertebrate soft tissues ($n = 9$), invertebrate biomineralized tissues ($n = 9$), and non-biomineralized tissues ($n = 11$). This ratio and the data shown in Fig. 2B characterize the degree to which organic matter is altered (in vivo or diagenetically) and identify glycoxidation or lipoxidation as the process responsible based on tissue-specific in vivo abundance of lipids and sugars as RCS sources. PFPs, protein fossilization products.

Tissue type	Composition	Remodeling	$\dfrac{[-\mathbf{C}-\mathbf{S}-]}{[-\mathbf{C}-\mathbf{N}-]}$	Alteration	RCS generation	Phylogenetic signal
Eggshell	Proteins, sugars	None	0.52	Low	Glycoxidation	High
Teeth	Proteins, lipids, sugars	Remodeled	1.40	High	Lipoxidation/ glycoxidation	Low
Bone	Proteins, sugars, lipids	Remodeled	0.98	High	Lipoxidation/ glycoxidation	Low
Vertebrate soft tissue	Proteins, lipids	None	0.22	Low	Glycoxidation	High
Invertebrate biomineralized tissue	Sugars, proteins, lipids	None	0.42	Low	Glycoxidation/ lipoxidation	High
Invertebrate soft tissue	Sugars, proteins	None	0.11	High	Glycoxidation	High

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

The degree of protein transformation in fossil metazoans 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_p); Fig. 2B] (*12*). Advanced cross-linking tends to produce heterocycles through condensation reactions, which yield characteristic *cis*-amide (A_d) 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^{-1}, representing *trans*-amides, and 1702 cm^{-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_p/A_d) >1 indicate that more peptide bonds are intact than are diagenetically 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 metazoans ($n = 76$) 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 metazoans ($n = 76$) 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^{-1} (S-heterocycles) and 1580 cm^{-1} (N-heterocycles) Raman shifts were determined for samples of every tissue category in the data matrix generated for the
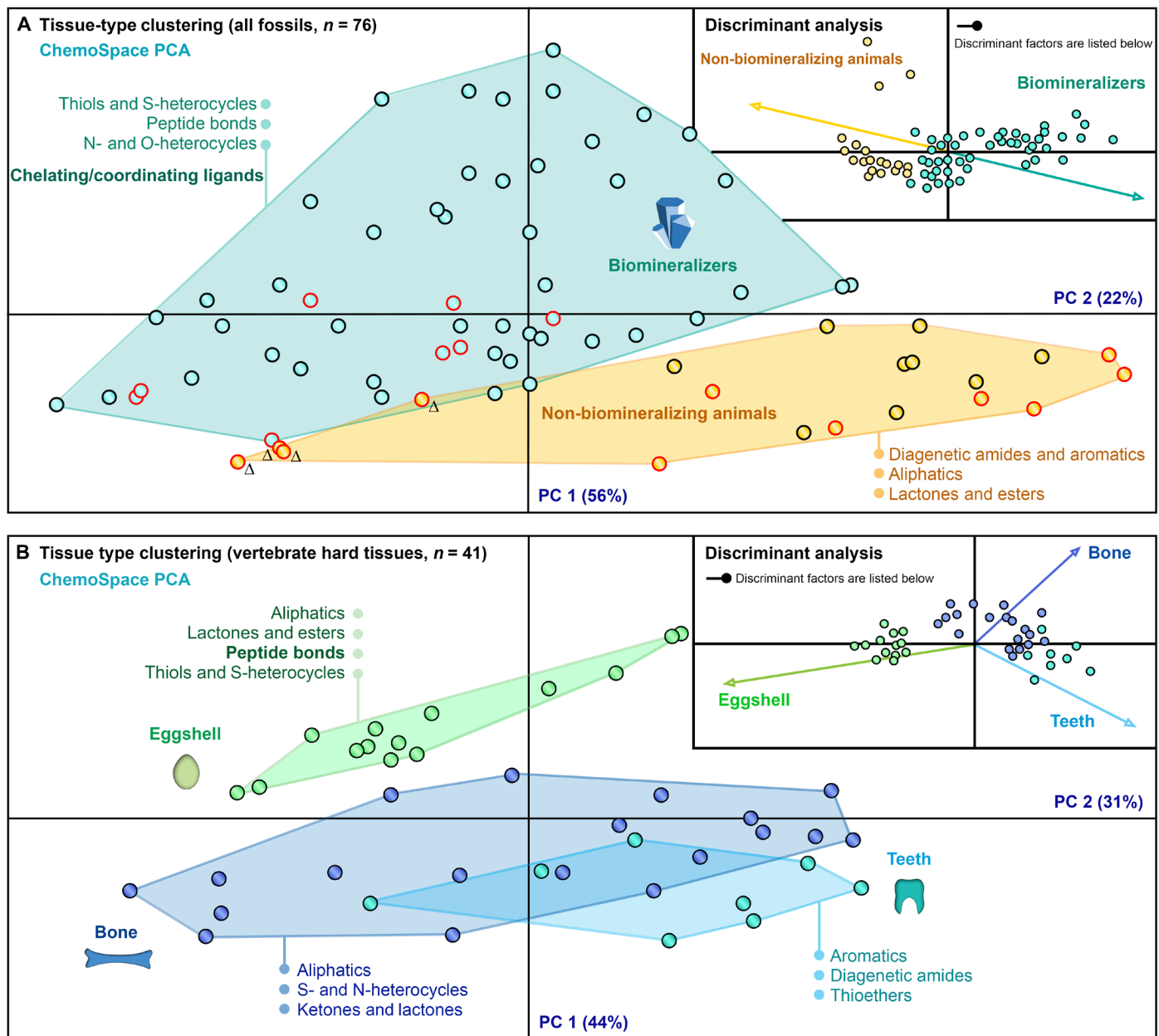
Fig. 3. Clustering of fossil animal tissues. (**A**) Biomineralization signal for all metazoan fossils based on spectral intensities at 24 Raman band positions ($n = 76$) (see Materials and Methods) complemented by a discriminant analysis of the biomineralized ($n = 56$, blue dots) and non-biomineralized ($n = 20$, 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); Δ, thermally matured samples (Cambrian Burgess Shale); blue arrow, trajectory of biomineralized tissues; yellow arrow, that of no-nbiomineralized 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.

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 [—C—S—]/[—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

types of cluster analysis: a cross-tissue analysis and individual cluster analyses for each tissue type. For the cross-tissue cluster analysis, all prescreened samples suitable for phylogenetic analyses were used, except for vertebrate eggshells, the siliceous sponge, and *Lithostrotion*, which represent tissues that contain proteins that act primarily as biomineral chelates and templates but lack any conserved proteinaceous structures. The resulting dataset of fossil metazoan tissues ($n = 27$) was subjected to a hierarchical cluster analysis in PAST 3 (Fig. 4A).

Tissue-specific taxon-character matrices were treated as separate categories and were subjected to hierarchical cluster analyses in PAST 3. A rooted topology was generated for each category (one way, rho). The resulting topologies were compared with published consensus trees (Fig. 4, A and C to E), and a measure of correspondence was calculated by assessing each node for correct resolution. Consensus trees are based on the published literature (*18–22*).

Using the taxon-character matrix (Supplementary Data, band set 4) for fossil bone tissues, a ChemoSpace analysis (Fig. 4B) was run in PAST 3. On the basis of the literature on fossil metabolic rates, data points were colored according to their metabolic rate. S-/N-heterocycle ratios were averaged for all categories (Table 1) to investigate why the accuracy of the phylogenetic signal differs between tissues.

RESULTS

Chemical transformations of metazoan biomolecules during fossilization

We compared both the molecular composition and functional diversity of organic matter in extant and fossil samples (Fig. 1). Spectra were grouped in tissue categories—teeth, bones, vertebrate soft tissues, biomineralized invertebrate samples, and non-biomineralized invertebrate samples—and modern and fossil samples were treated separately. Unaltered structural biomolecules in extant samples are characterized by a high diversity of organic functions (*z* axis in Fig. 1A) and cover a distinct area {*x,y*} in the ChemoSpace (PC 1, 63%, *x* axis; PC 2, 14%, *y* axis; Supplementary Data, 3D ChemoSpace). The equivalent organic phase in fossils, in contrast, showed a substantial decrease in functional diversity and occupies a partially overlapping but largely distinct {*x,y*} area (Fig. 1A). Modern, unaltered biomolecules form an almost {*x,y*}-parallel plateau of high functional diversity and cover a different {*x,y*} area than the fossil samples in this 3D ChemoSpace. During fossilization, the total organic tissue composition converges on a functionally simpler composition, as the diversity of distinct organic groups falls to a third in fossil tissues (Fig. 1A, compare fig. S1). The shift in the tissue composition during fossilization represents an increase in glycoxidation and lipoxidation markers (Fig. 1A, red eigenvector). Modern, remodeled, and vascularized tissues, such as teeth (dentine) and bones, are known to record cross-links formed in vivo resulting from metabolic stress (*23–26*), resulting in a shift toward increased amounts of AGEs and ALEs in the ChemoSpace. A discriminant analysis (Fig. 1A) performed on the character matrix for fresh and fossil tissues revealed that the transformation of thiols into thioethers/S-heterocycles, and amines into N-heterocycles are key distinctions between fresh and fossil tissues (section S4 and fig. S1). The loss of functional diversity and the change in the portion of ChemoSpace occupied indicate substantial chemical transformation of organic matter during fossilization.

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, carbonyls, 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
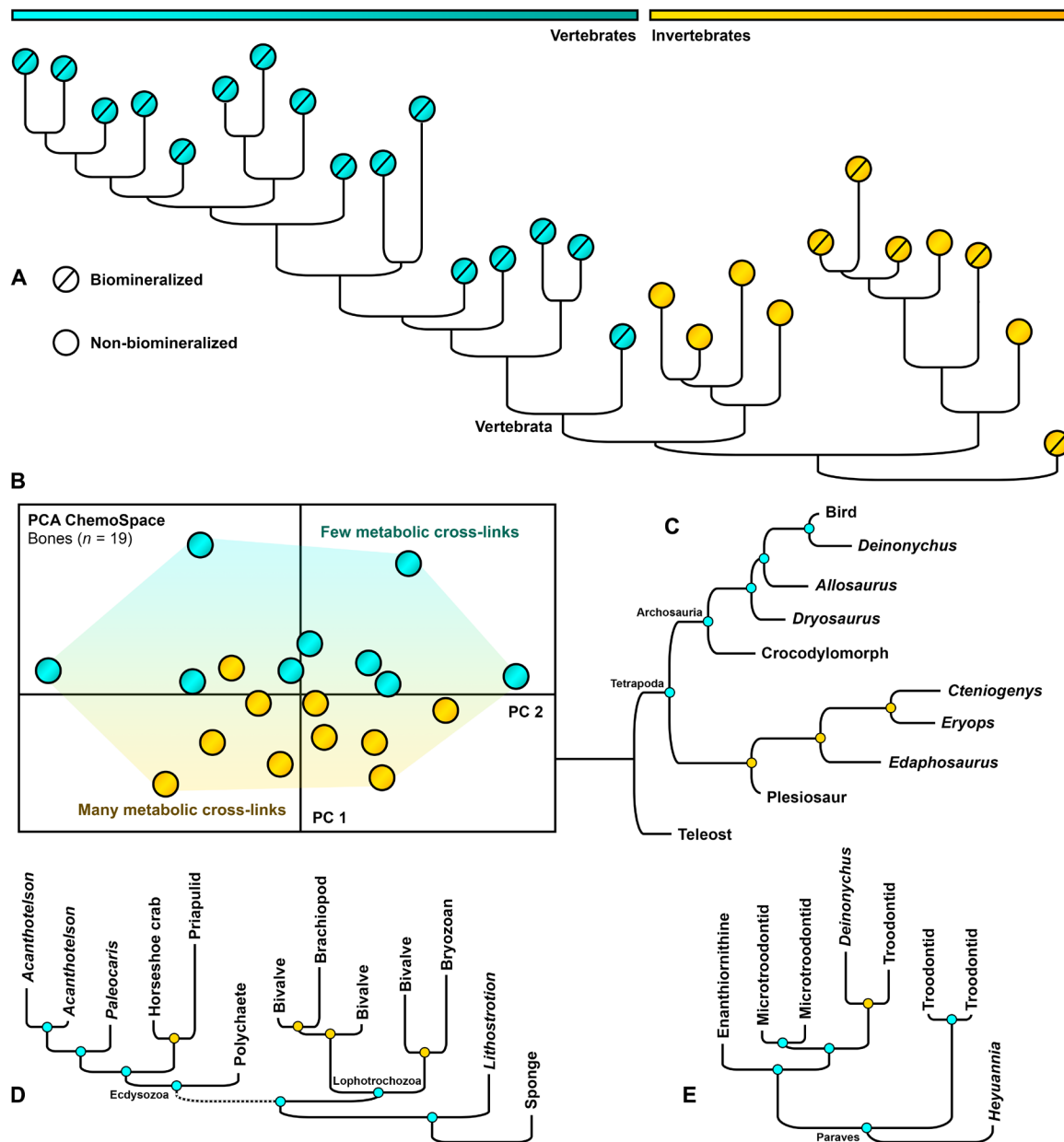
Fig. 4. Assessment of the phylogenetic signal preserved in fossils. Hierarchical cluster analysis of metazoan fossils based on Raman intensities at 20 band positions characterizing the specific composition of N– and S–cross-links ($n = 37$); vertebrate fossil soft tissues are not included (see Materials and Methods). Samples in each cluster are selected by tissue type. Node colors: blue, correct; orange, incorrect solution (Supplementary Data, Specimen Data). The topology calculated from fossil tooth spectra can be found in the Supplementary Data. (**A**) Cross-tissue–type cluster analysis of phylogenetic signals (Materials and Methods). Orange dots, invertebrates; blue dots, vertebrates. (**B**) Vascularized bone ($n = 19$, samples as in Fig. 3B) in a PCA ChemoSpace based on the abundance of in vivo metabolic cross-links. Blue dots, metabolic rate <1 ml O_2 hour^{-1} g$^{-0.67}$; orange dots, metabolic rate >1 ml O_2 hour^{-1} g$^{-0.67}$. (**C**) Fossil bones: Correspondence with published consensus trees is 62%. (**D**) Biomineralized and non-biomineralized invertebrate tissues (clusters determined independently and combined): Correspondence is 75% and 40%, respectively; even derived nodes are accurately resolved (Crustacea in Ecdysozoa). (**E**) Fossil eggshells: Correspondence is 83%.

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 biominerals also differ in the content 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 biominerals (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 $[(C{-}S)/(C{-}N)]$ ratio 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*, to guarantee comparability of the protein-derived heterogeneities (Fig. 4A). In this cross-tissue analysis, vertebrates fall out as a monophyletic group nested within invertebrates.

A ChemoSpace of fossil bones (Fig. 2B) reveals the presence of a metabolic signal in the relative abundance of S- and N-heterocycles in vascularized tissues (compare to $[C{-}S]/[C{-}N]$ ratios in Table 1), likely overprinting the weaker phylogenetic signal (12, 23–26). The $[{-}C{-}S{-}]/[{-}C{-}N{-}]$ ratios for fossil tissues with a strong phylogenetic signal range from 0.11 to 0.52, but ratios for bones and teeth (0.98 to 1.40), the only vascular tissues in the dataset, indicate increased accumulation of S-heterocycles. Since vascular tissues are remodeled, and thereby directly affected by metabolism, phylogenetic signals in their structural proteins may be overprinted in vivo by a metabolic signal (11, 25, 26). Taxa with few in vivo metabolic cross-links yield positive values for PC 1 in Fig. 4B, while taxa with high metabolic rates yield negative values.

Nodes in the topologies of individual tissue types (Fig. 4, C to E) were assessed for expected or incorrect resolution [based on (18–22); Fig. 4]. Organic matter in fossil eggshell (83% correspondence with expected topology), non-biomineralized invertebrate tissues (75%), bone (63%), biomineralized invertebrate tissues (40%), and teeth (25%) preserves phylogenetic signals of different quality (Fig. 4, B to D). However, most fossil tissues allow the reliable reconstruction of higher-rank metazoan relationships. These results reflect the differing degrees of diagenetic alteration across tissue categories as revealed by the cis-/trans-amide ratios (Fig. 2A), the ChemoSpace PCAs (Fig. 3), and the ratios of S- and N-heterocycles (Table 1), which correlate in vascularized tissues with metabolic capabilities.

DISCUSSION

The analyses presented here show that biomolecule fossilization products associated with fossil metazoans are endogenous and readily distinguished from organics in the associated sediments on the basis of preserved peptide bonds, thioethers, and S-heterocycles (Fig. 2D). These compositional differences (Fig. 2D) between metazoan and sedimentary fossil organic matter are also evident when fossil stromatolites and plants are included in the sample suite (27). A fossil-sediment ChemoSpace (e.g., Fig. 2D), cis-/trans-amide ratios, and the metazoan biomarkers (abundant thioethers, S-heterocycles, and trans-amides) offer a threefold approach to determine the utility of specimens for future investigations.

Peptide bonds (trans-amides) survive to different degrees in different tissues (Fig. 2B and section S8). Thioethers and S-heterocycles, on the other hand, form mostly de novo through 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 thioethers, 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
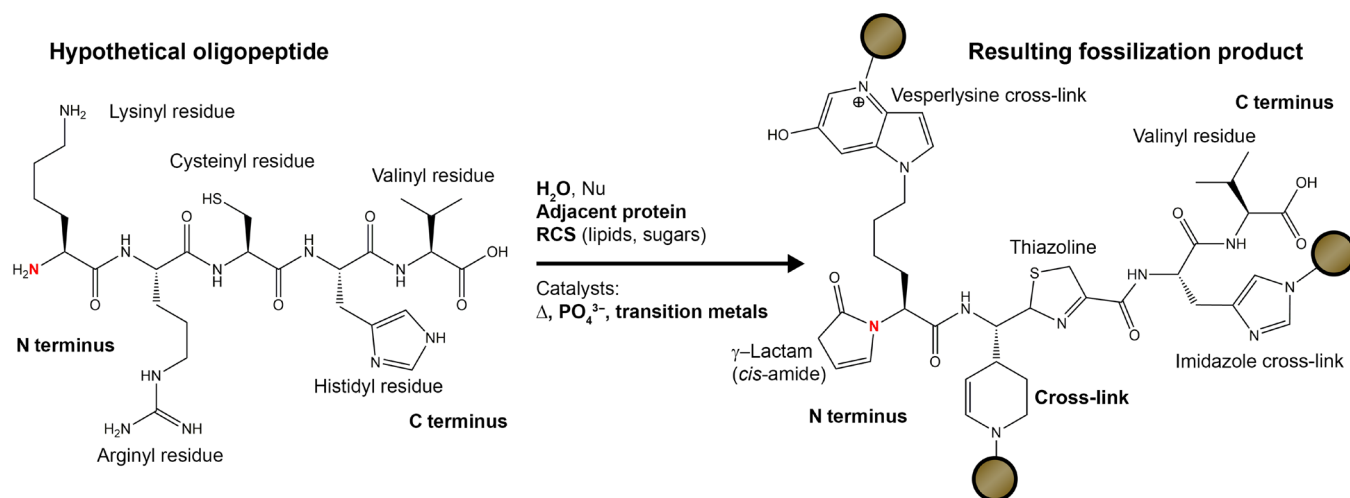
Fig. 5. Molecular mechanism for the preservation of phylogenetic signals in fossils (based on data in Figs. 1A, 2A, 3, and 4, A and C to E). The reactive amino acids lysine, arginine, histidine, cysteine, and the protein N terminus (labeled in red) form characteristic fossilization products under suitable conditions. Brown circles indicate extensions of the geopolymer. Nu, nucleophile; Δ, thermal energy input.

hard tissues (*9*) but also plays a universal role in the preservation of soft tissues in deep time. The chemical transformation of proteins, lipids, and sugars into more stable N-, O-, and S-heterocycles is favored under certain conditions, such as water saturation, an oxidative chemomilieu, and the presence of catalyzing transition metals or phosphate (*9*) [reviewed in (*12*)]. Integrating data on biological signals in fossil organic matter with known glycoxidation and lipoxidation reaction schemes shows that, depending on the original tissue composition, either sugars or lipids represent the major source of RCS (Fig. 1 and fig. S1) (*12*, *33*). RCS are known to target amino acid residues of structural proteins, particularly lysinyl, arginyl, histidyl, and cysteinyl residues (Fig. 5; see section S7 for definitions) (*12*, *26*, *33*). Other amino acids are less likely to react in the initial stages of the oxidative cross-linking process (*12*, *26*, *33*) and are therefore detected more commonly in ancient materials (*30*). Amine-bearing residues (lysinyl, arginyl, and histidyl residues) generate N-heterocycles, whereas cysteinyl residues, which contain a characteristic thiol group, generate thioethers and S-heterocycles (Fig. 1 and fig. S1) (*12*, *26*, *33*). Amine-bearing residues cross-link with RCS of both lipid and sugar origin, whereas cysteinyl residues cross-link selectively in the presence of lipid-derived RCS (Fig. 2C) (*12*, *26*, *33*). The proportion of amine- and thiol-bearing amino acid residues in a protein, and of N- and S-heterocycles in its fossil derivative, reflects the original amino acid composition (Figs. 1A and 2C and fig. S1). Detailed characterization of the range of protein-RCS cross-linked products formed during fossilization will open up previously unidentified applications of biological signals from fossils (*34*, *35*).

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

View/request a protocol for this paper from *Bio-protocol*.

REFERENCES AND NOTES

1. J. F. Bone, *Animal anatomy and physiology* (Reston Publishing Company, 1979).
2. J. Black, G. Hastings, *Handbook of biomaterial properties* (Springer US, 2013).
3. F. Crick, Central dogma of molecular biology. *Nature* **227**, 561–563 (1970).
4. L. Addadi, S. Weiner, Interactions between acidic proteins and crystals: Stereochemical requirements in biomineralization. *Proc. Natl. Acad. Sci. U.S.A.* **82**, 4110–4114 (1985).
5. D. E. G. Briggs, R. E. Summons, Ancient biomolecules: Their origins, fossilization, and role in revealing the history of life. *Bioessays* **36**, 482–490 (2014).
6. M. Buckley, A. Walker, S. Y. W. Ho, Y. Yang, C. Smith, P. Ashton, J. T. Oates, E. Cappellini, H. Koon, K. Penkman, B. Elsworth, D. Ashford, C. Solazzo, P. Andrews, J. Strahler, B. Shapiro, P. Ostrom, H. Gandhi, W. Miller, B. Raney, M. I. Zylber, M. T. P. Gilbert, R. V. Prigodich, M. Ryan, K. F. Rijsdijk, A. Janoo, M. J. Collins, Weighing the mass spectrometric evidence for authentic *Tyrannosaurus rex* collagen. *Science* **319**, 33 (2008).
7. M. Bern, B. S. Phinney, D. Goldberg, Reanalysis of *Tyrannosaurus rex* mass spectra. *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. Jersie-Christensen, J. V. Olsen, J. Haile, 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. Kwekason, 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).
9. J. Wiemann, M. Fabbri, T.-R. Yang, K. Stein, P. M. Sander, M. A. Norell, D. E. G. Briggs, Fossilization transforms vertebrate hard tissue proteins into N-heterocyclic polymers. *Nat. Commun.* **9**, 4741 (2018).
10. P. Adam, P. Schneckenburger, P. Schaeffer, P. Albrecht, Clues to early diagenetic sulfurization processes from mild chemical cleavage of labile sulfur-rich geomacromolecules. *Geochim. Cosmochim. Acta* **64**, 3485–3503 (2000).
11. R. P. Evershed, H. A. Bland, P. F. van Bergen, J. F. Carter, M. C. Horton, P. A. Rowley-Conwy, Volatile compounds in archaeological plant remains and the Maillard reaction during decay of organic matter. *Science* **278**, 432–433 (1997).
12. G. Vistoli, 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 of formation. *Free Radic. Res.* **47**, 3–27 (2013).

13. J. Wiemann, T.-R. Yang, M. A. Norell, Dinosaur egg colour had a single evolutionary origin. *Nature* **563**, 555–558 (2018).

14. E. M. Boatman, M. B. Goodwin, H.-Y. N. Holman, S. Fakra, W. Zheng, R. Gronsky, M. H. Schweitzer, Mechanisms of soft tissue and protein preservation in *Tyrannosaurus rex. Sci. Rep.* **9**, 15678 (2019).

15. A. M. Pawlak, J. R. Beattie, J. V. Glenn, A. W. Stitt, J. J. McGarvey, Raman spectroscopy of advanced glycation end products (AGEs), possible markers for progressive retinal dysfunction. *J. Raman Spectrosc.* **39**, 1635–1642 (2008).

16. H. Salehi, E. Terrer, I. Panayotov, B. Levallois, B. Jacquot, H. Tassery, F. Cuisinier, Functional mapping of human sound and carious enamel and dentin with Raman spectroscopy. *J. Biophotonics* **6**, 765–774 (2013).

17. P. D. Lewis, G. E. Menzies, Vibrational spectra, principal components analysis and the horseshoe effect. *Vib. Spectrosc.* **81**, 62–67 (2015).

18. T. R. Simões, M. W. Caldwell, M. Tałanda, M. Bernardi, A. Palci, O. Vernygora, F. Bernardini, L. Mancini, R. L. Nydam, The origin of squamates revealed by a Middle Triassic lizard from the Italian Alps. *Nature* **557**, 706–709 (2018).

19. A. Meyer, R. Zardoya, Recent advances in the (molecular) phylogeny of vertebrates. *Annu. Rev. Ecol. Evol. Syst.* **34**, 311–338 (2003).

20. S. L. Brusatte, G. T. Lloyd, S. C. Wang, M. A. Norell, Gradual assembly of avian body plan culminated in rapid rates of evolution across the dinosaur-bird transition. *Curr. Biol.* **24**, 2386–2392 (2014).

21. H. Glenner, A. J. Hansen, M. V. Sørensen, F. Ronquist, J. P. Huelsenbeck, E. Willerslev, Bayesian inference of the metazoan phylogeny: A combined molecular and morphological approach. *Curr. Biol.* **14**, 1644–1649 (2004).

22. J. Gauthier, Saurischian monophyly and the origin of birds. *Mem. Calif. Acad. Sci.* **8**, 1–55 (1986).

23. J. A. Hardin, N. Cobelli, L. Santambrogio, Consequences of metabolic and oxidative modifications of cartilage tissue. *Nat. Rev. Rheumatol.* **11**, 521–529 (2015).

24. E. D. Schleicher, E. Wagner, A. G. Nerlich, Increased accumulation of the glycoxidation product N(epsilon)-(carboxymethyl)lysine in human tissues in diabetes and aging. *J. Clin. Invest.* **99**, 457–468 (1997).

25. N. Verzijl, J. De Groot, E. Oldehinkel, R. A. Bank, S. R. Thorpe, J. W. Baynes, M. T. Bayliss, J. W. Bijlsma, F. P. Lafeber, J. M. Tekoppele, Age-related accumulation of Maillard reaction products in human articular cartilage collagen. *Biochem. J.* **350**, 381–387 (2000).

26. J. M. Silván, J. van de Lagemaat, A. Olano, M. D. Del Castillo, Analysis and biological properties of amino acid derivates formed by Maillard reaction in foods. *J. Pharm. Biomed. Anal.* **41**, 1543–1551 (2006).

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).

28. D. Gournis, M. A. Karakassides, D. Petridis, Formation of hydroxyl radicals catalyzed by clay surfaces. *Phys. Chem. Miner.* **29**, 155–158 (2002).

29. N. S. Gupta, O. Cambra-Moo, D. E. G. Briggs, G. D. Love, M. A. Fregenal-Martinez, R. E. Summons, Molecular taphonomy of macrofossils from the Cretaceous Las Hoyas Formation, Spain. *Cretac. Res.* **29**, 1–8 (2008).

30. A. Kemp, Amino acid residues in conodont elements. *J. Pal.* **76**, 518–528 (2002).

31. J. Alleon, S. Bernard, C. Le Guillou, D. Daval, F. Skouri-Panet, M. Kuga, F. Robert, Organic molecular heterogeneities can withstand diagenesis. *Sci. Rep.* **7**, 1508 (2017).

32. D. E. G. Briggs, R. P. Evershed, M. J. Lockheart, The biomolecular paleontology of continental fossils. *Paleobiology* **26**, 169–193 (2000).

33. F. J. Hidalgo, M. Alaiz, R. Zamora, Effect of pH and temperature on comparative nonenzymatic browning of proteins produced by oxidized lipids and carbohydrates. *J. Agr. Food Chem.* **47**, 742–747 (1999).

34. V. E. McCoy, J. Wiemann, J. C. Lamsdell, C. D. Whalen, S. Lidgard, P. Mayer, H. Petermann, D. E. G. Briggs, Chemical signatures of soft tissues distinguish between vertebrates and invertebrates from the Carboniferous Mazon Creek Lagerstätte of Illinois. *Geobiology* **2020**, 1–6 (2020).

35. M. A. Norell, J. Wiemann, M. Fabbri, C. Yu, C. A. Marsicano, A. Moore-Nall, D. J. Varricchio, D. Pol, D. K. Zelenitsky, The first dinosaur egg was soft. *Nature* (2020); doi.org/10.1038/s41586-020-2412-8.

Citation: J. Wiemann, J. M. Crawford, D. E. G. Briggs, Phylogenetic and physiological signals in metazoan fossil biomolecules. *Sci. Adv.* **6**, eaba6883 (2020).