Response to Comment on "Protein Sequences from Mastodon and Tyrannosaurus rex Revealed by Mass Spectrometry"

John M. Asara, *et al.*

Science **319**, 33d (2008);
DOI: 10.1126/science.1147364

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005.

Response to Comment on "Protein Sequences from Mastodon and *Tyrannosaurus rex* Revealed by Mass Spectrometry"

John M. Asara[1,2]* and Mary H. Schweitzer[3,4]

We sequenced six endogenous collagen peptides from *Tyrannosaurus rex* bone fragments using mass spectrometry. Five sequences match birds, but only two match amphibians, supporting dinosaur-bird relationships. Buckley *et al.* reinterpret and misinterpret our data and question sequence authenticity, but they used a suboptimal phylogenetic algorithm to analyze only a subset of reported sequences and they suggest analyses that are less sensitive and less specific than mass spectrometry. We disagree and use data to explain.

Buckley *et al*. (*1*) use authentication tests developed for ancient DNA to question the authenticity of peptide sequences we obtained from an exceptionally well-preserved 68-million-year-old *Tyrannosaurus rex* fossil (*2*, *3*). We disagree with their assessment for the reasons explained below. We first wish to point out that Table 1 lists the six high-confidence peptide sequences for the exceptionally well-preserved 68-million-year-old *T. rex* fossil (MOR 1125) acquired by tandem mass spectrometry (MS/MS) (*3*). All future phylogenetic analyses or critiques should be performed using these sequences.

Turning to the issue of collagen survival brought up by Buckley *et al*. (*1*), the preservational state of *T. rex* specimen MOR 1125 was described previously (*4*) and showed unexpected retention of soft tissue and cellular structures. Its burial under approximately 1000 m^3 of semiconsolidated sandstone limited its exposure to modern environmental influences (water, light, and air), contributing to the retention of endogenous protein identified first by immunochemical assays, then by ion trap mass spectrometry (*2*, *3*). Sequences were derived from multiple extractions conducted over about 1.5 years. It has been reported that cross-linking (i.e., packing) of collagen molecules increases their thermal stability, which contrasts with statements in (*1*) and which, in combination with association to bone mineral, contributes to the longevity of the molecule (*5*). Combined with supporting data, including in situ immunochemistry and enzyme-linked immunosorbent assays, atomic force microscopy, in situ time-of-flight secondary ion mass spectrometry, and various microscopic analyses (*6*), the sequence data suggest a reevaluation of theories in the Buckley *et al*. comment regarding protein longevity and preservation.

Regarding the issue of contamination risk, we showed that endogenous protein fragments derived from dinosaur bone extracts were in low/sub femtomole amounts (*2*) from milligrams of extract. Because of these low concentrations, contamination from soil bacteria and human keratin (common in most samples from dust, skin, and hair) were present. Contamination with these common sources would make bulk amino acid analysis (AAA) data meaningless, as AAA is not protein sequence specific to address the source of the amino acids. We were able to sequence the keratin and bacterial peptides separately from collagen using microcapillary liquid chromatography (LC) MS/MS. Collagen type I, concentrated in bone, is not a common source of protein contamination, as Buckley *et al*. note. Not all *T. rex* peptides are consistent with a common organism, so multiple separate contamination events are required to support the Buckley *et al*. argument for contamination as a source. This is not supported by our consistently negative controls. Furthermore, the sequences shown in Table 1 do not support links to amphibians, and amphibians are neither native to nor present at Hell Creek, nor have they ever been present in either of the labs where analysis was performed. Bone from amphibians or related organisms are not a logical source of modern contamination. Given that we observed not only gross morphological preservation but also apparent preservation of other characteristics such as elasticity and banding, in situ localization of collagen antibodies to tissues with negative controls, and collagen-consistent sequence data consistent with birds, the most parsimonious explanation is that the material is endogenous. To challenge this hypothesis requires data that support alternative hypotheses internally consistent for all data presented, not a small subset of data as in the comment by Buckley *et al*. (*1*).

Buckley *et al*. conduct a phylogenetic analysis on a small subset of sequence data that we produced but that they inappropriately reinterpret. They included only four sequences for mastodon and six for ostrich, whereas we published 78 for mastodon and 93 for ostrich (*3*). Buckley *et al*. claim that several spectra cannot be interpreted, although the fragmentation patterns and signal-to-noise-ratios are of interpretable quality. They included only five of the six peptides that we identified in *T. rex* extracts. Specifically, they chose to omit the collagen type I peptide sequence GLPGESGAVGPAGPIGSR from the α2 chain, without clear justification for doing so. This omission is important, because it is a unique match to chicken (birds). They use a single neighbor-joining algorithm to generate a phylogenetic hypothesis of protein source and do not

Table 1. Clarified 68-million-year-old *T. rex* collagen peptide sequences. Hydroxylation sites are noted with an asterisk (*) immediately following the modified residue. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

Peptide sequence	Protein	Organism identity
GATGAP*GIAGAP*GFP*GAR	Collagen α1t1	Identical to chicken and frog (amphibians)
GSAGPP*GATGFP*GAAGR	Collagen α1t1	Identical to multiple mammals, including chicken (no amphibian identity)
GVQGPP*GPQGPR	Collagen α1t1	Identical to chicken and opossum (no amphibian identity)
GLPGESGAVGPAGPIGSR	Collagen α2t1	Identical to chicken only (no amphibian identity)
GVVGLP*GQR	Collagen α1t1	Identical to multiple organisms, including chicken and newt (amphibians)
GAPGPQGPSGAP*GPK	Collagen α1t1	Unique *T. rex* sequence (no amphibian or bird identity)

[1]Division of Signal Transduction, Beth Israel Deaconess Medical Center, Boston, MA 02115, USA. [2]Department of Pathology, Harvard Medical School, Boston, MA 02115, USA. [3]Department of Marine, Earth, and Atmospheric Sciences, North Carolina State University, Raleigh, NC 27695, USA. [4]North Carolina Museum of Natural Sciences, Raleigh, NC 27601, USA.

*To whom correspondence should be addressed. E-mail: jasara@bidmc.harvard.edu

validate their hypothesis with other algorithms. It is commonly agreed by evolutionary biologists that the preferred practice is to use multiple methods to generate phylogenetic hypotheses and then look for agreement. For example, use of Bayesian analysis, parsimony, and likelihood in addition to neighbor-joining are important to validate the results from a single algorithm (*7*). Use of all of our data would make it impossible to group *T. rex* with amphibians. The sequence clustering of the majority of the dinosaur sequences (five of six) with birds (chicken) and not amphibians (only two of six) based on our data invalidates the phylogenetic conclusions proposed by Buckley *et al*. and upholds our original conclusions that birds are the closest living organisms to dinosaurs among all organisms for which collagen sequence data are available. We agree that only six high-confidence collagen-derived peptides recovered from *T. rex* may not be enough to generate robustly supported phylogenetic trees; however, the amino acid sequences of the peptides are sufficient to identify the source protein, collagen type I, and to assess dinosaur peptide sequence similarity/identity to other organisms using BLAST analysis. Future phylogenetic analyses using all of our data, in addition to new protein sequence data for critical taxa (birds and reptiles), are needed to make an accurate assessment of *T. rex* evolutionary placement.

Buckley *et al.* make five challenges to our report of endogenous collagen protein within dinosaur bone: (i) Although no standards for replication and authentication of protein data currently exist for ancient protein data, we did not apply standards currently recommended for DNA studies. (ii) We did not have our extractions conducted in multiple labs. (iii) We did not perform inhibition studies in immunohistochemical assays using multiple taxa. (iv) No standard amino acid analyses were conducted. (v) We did not account for common deamidation modifications. Our responses to these criticisms are as follows.

First, as Buckley *et al*. observe, no standards exist for ancient protein analyses. Because we are not enzymatically amplifying molecules, we are unlikely to observe exogenous contaminants in our many analyses, including localization of antibody-antigen complexes and collagen-consistent amino acid residues to the samples. No collagen signal above background was observed in multiple controls conducted in tandem in any of these diverse assays. Second, few labs exist that are dedicated to ancient protein analyses, and labs routinely capable of performing these assays generally conduct similar analyses on modern protein samples; hence, induction of modern contaminating molecules are much more likely from nondedicated protein labs. In addition, very little proteinaceous material was available for analysis, and this diminished over the course of our analyses as the material responded to modern environmental conditions. However, provided that the amount of sequenceable material is sufficient, we intend to send future samples to other labs for verification. The choice between potentially acquiring more sequences versus the risk of sending the extract to another lab for confirmation was obvious. Third, although we did not conduct multitaxon inhibition studies, we did perform these inhibition assays using bird collagen, which is more relevant and appropriate and has the strongest inhibition capacity for antibodies to bird collagen used in this study. Fourth, concentrations of proteinaceous material were insufficient to conduct amino acid analysis. Bulk amino acid analyses do not contribute enough useful information (in terms of assigning resulting signal to source protein). However, we were able to localize amino acids consistent with collagen to dinosaur tissues only and not to embedding material or analytical mounting materials. Controls, including those conducted on surrounding sandstones, undertaken in tandem with samples for each assay, were consistently negative (*6*). Fifth, we note that all instances of glutamic acid (E) and aspartic acid (D) that occur in sequences are not the result of deamidation events but may be the result of translations from genomic coding sequence. The ion trap mass spectrometer used in the study, although extremely sensitive, is capable of relatively low resolution and low mass accuracy, so these modification events were difficult to detect, though possibly present. Analysis of other fossil specimens using our newly acquired ultrahigh-resolution and mass accuracy Orbitrap mass spectrometer clearly show deamidation events. It is difficult to distinguish between posttranslational modifications and diagenetic deamidation because rates of occurrences of these modifications under natural conditions have not been well studied and because the *T. rex* and mastodon sequences were not analyzed before fossilization. Because deamidation occurs on freshly prepared proteins in our lab at levels similar to many "fossilized" proteins, we question whether deamidation is a particularly useful marker for age.

We continue our analyses of MOR 1125 and other dinosaur specimens not only to verify our original results but also to demonstrate preservation and molecular recovery in other specimens. We also continue to develop new, more sensitive, and more accurate methodologies using new instrumentation. We presented our analytical data as a first step in understanding protein preservation and evolutionary comparisons for extinct taxa. Although more sequence data from multimillion-year-old ancient fossils is a primary goal, the small sampling of data that we have shown demonstrates the possibility of obtaining multimillion-year-old sequences using mass spectrometry. Our papers provide a baseline methodology for determining sequences from ancient samples while pushing the limits of state-of-the-art instrumental sensitivity. Further work, based on experience and experimental data, will contribute to establishing standards for analyzing ancient protein samples.

References and Notes

1. M. Buckley *et al*., *Science* **319**, 33 (2008); www.sciencemag.org/cgi/content/full/319/5859/33c.
2. J. M. Asara, M. H. Schweitzer, L. M. Freimark, M. Phillips, L. C. Cantley, *Science* **316**, 280 (2007).
3. J. M. Asara *et al*., *Science* **317**, 1324 (2007).
4. M. H. Schweitzer, J. L. Wittmeyer, J. R. Horner, J. K. Toporski, *Science* **307**, 1952 (2005).
5. C. A. Miles, M. Ghelashvili, *Biophys. J.* **76**, 3243 (1999).
6. M. H. Schweitzer *et al*., *Science* **316**, 277 (2007).
7. J. J. Wiens, *Syst. Biol.* **52**, 528 (2003).
8. We thank C. L. Organ, L. M. Freimark, M. Phillips, L. C. Cantley, and J. R. Horner for helpful discussions.

10 July 2007; accepted 30 November 2007
10.1126/science.1147364