original protein molecules became irreversibly cross-linked, both to similar molecules and to mineral or exogenous organic components. These cross-linking reactions may have been initiated by unstable metal ions that formed free radicals (*30*, *31*), which in turn reacted with organic molecules to form polymers (*6*, *7*, *9*, *32*). We propose that the unstable metal ions were derived from the post mortem degradation of iron-containing dinosaur biomolecules such as hemoglobin, myoglobin, and possibly cytochromes (*9*, *31*). Once stabilized by these cross-linking reactions, the molecules were no longer available as substrates for further degradative reactions.

The intimate relationship between apatite and the organic phase of bone also contributes to the preservation of organic matter (*16*, *33*–*38*), but we propose that the mineral phase may be stabilized by this relationship as well. The presence of biogenic apatite in these 68-million-year-old bones can only be rationalized by protection from an intact organic phase, which in turn is only satisfied by a synergistic relationship between collagen and mineral phases. Whereas extant bone retains no detectable calcium after days to weeks of demineralization, dinosaur bone retains a fraction of recognizable apatite crystals after months of treatment (fig. S1). Another contributing factor in the retention of original mineral may be that apatite is stabilized in the presence of calcite (*33*). Sandstones surrounding MOR 1125 contain abundant calcite cements.

The depositional environments may affect organic preservation in other ways. Comparison of fossils from a variety of environments indicates that those derived from sandstones are more likely to retain soft tissues and/or cells (*9*). We hypothesize that the porosity of sandstones may facilitate draining of enzymes of decay and suppurating fluids as the organism degrades, whereas organisms buried in nonporous mudstones or clays may be exposed to these longer and therefore may be more completely degraded.

Our findings indicate the need for optimizing methods of extraction and handling of fossil material. In particular, the decrease in signal we observed over time supports the need to establish field collection and storage of fossils according to protocols that allow future analytical studies (*29*).

The data presented here illustrate the value of a multidisciplinary approach to the characterization of very old fossil material and validate sequence data reported elsewhere (*10*). The inclusion of fossil-derived molecular sequences into existing phylogenies may provide greater resolution and may allow reconstruction of character evolution beyond what is currently possible. Elucidating modifications to ancient molecules may shed light on patterns of degradation and diagenesis. The presence of original molecular components is not predicted for fossils older than a million years (*1*–*7*), and the discovery of collagen in this well-preserved dinosaur supports the use of actualistic conditions to formulate molecular degradation rates and models, rather than relying on theoretical or experimental extrapolations derived from conditions that do not occur in nature.

### References and Notes


1. T. Lindahl, *Nature* **365**, 700 (1993).
2. G. Eglinton, G. A. Logan, *Philos. Trans. R. Soc. London Ser. B* **333**, 315 (1991).
3. J. L. Bada, X. Y. S. Wang, H. Hamilton, *Philos. Trans. R. Soc. London Ser. B* **354**, 77 (1999).
4. M. Hoss, *Nature* **404**, 453 (2000).
5. D. E. G. Briggs, R. P. Evershed, M. J. Lockheart, *Paleobiology* **26**, 169 (2000).
6. D. E. G. Briggs, *Annu. Rev. Earth Planet. Sci.* **31**, 275 (2003).
7. B. A. Stankiewicz *et al.*, *Geology* **28**, 559 (2000).
8. M. H. Schweitzer, J. L. Wittmeyer, J. R. Horner, J. B. Toporski, *Science* **307**, 1952 (2005).
9. M. H. Schweitzer, J. L. Wittmeyer, J. R. Horner, *Proc. R. Soc. London Ser. B* **274**, 183 (2007).
10. J. M. Asara, M. H. Schweitzer, L. M. Freimark, M. Phillips, L. C. Cantley, *Science* **316**, 280 (2007).
11. See supporting material on *Science* Online.
12. M. H. Schweitzer, J. L. Wittmeyer, J. R. Horner, *Science* **308**, 1456 (2005).
13. L. Zylberberg, *C. R. Palevol.* **3**, 591 (2004) and references therein.
14. M. van der Rest, *Bone* **3**, 187 (1991).
15. V. Ottani, M. Raspanti, A. Ruggeri, *Micron* **32**, 251 (2001).
16. S. Weiner, H. D. Wagner, *Annu. Rev. Mater. Sci.* **28**, 271 (1998).
17. G. Nemethy, H. A. Scherega, *Biochemistry* **25**, 3184 (1986).
18. J. G. Bann, H. P. Bachinger, *J. Biol. Chem.* **275**, 24466 (2000).
19. D. R. Keene, L. Y. Sakai, R. E. Burgeson, *J. Histochem. Cytochem.* **39**, 59 (1991).
20. N. Tuross, L. Stathoplos, *Methods Enzymol.* **224**, 121 (1993).
21. P. Semal, R. Orban, *J. Archaeol. Sci.* **22**, 463 (1995).
22. R. Avci *et al.*, *Langmuir* **21**, 3584 (2005).
23. K. Mosbach, *Trends Biochem. Sci.* **19**, 9 (1994).
24. F. A. Rimblot-Baly, A. de Ricqles, L. Zylberberg, *Ann. Paleontol.* **81**, 49 (1995).
25. J. M. Hughes, M. Cameron, K. D. Crowley, *Am. Mineral.* **74**, 870 (1989).
26. B.-A. Gotliv *et al.*, *J. Struct. Biol.* **156**, 320 (2006).
27. Accession number P02457, NCBI nonredundant protein database.
28. T. Holtz, *Gaia* **15**, 5 (1998) and references therein.
29. M. Pruvost *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 739 (2007).
30. J. M. C. Gutteridge, D. A. Rowley, B. Halliwell, *Biochem. J.* **199**, 263 (1981).
31. P. Mladěnka, T. Simunek, M. Hubl, R. Hrdina, *Free Radic. Res.* **40**, 263 (2006).
32. F. Q. Schafer, S. Yue Qian, G. R. Buettner, *Cell. Mol. Biol.* **46**, 657 (2000).
33. F. Berna, A. Matthews, S. Weiner, *J. Archaeol. Sci.* **31**, 867 (2004).
34. M. J. DeNiro, S. Weiner, *Geochim. Cosmochim. Acta* **52**, 2415 (1988).
35. M. J. Glimcher, L. Cohen-Solal, D. Kossiva, A. deRicqles, *Paleobiology* **16**, 219 (1990).
36. T. Schmidt-Schultz, M. Schultz, *Am. J. Phys. Anthropol.* **123**, 30 (2004).
37. G. A. Sykes, M. J. Collins, D. I. Walton, *Org. Geochem.* **23**, 1059 (1995).
38. M. Salamon, N. Tuross, B. Arensburg, S. Weiner, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 13783 (2005).


39. We thank J. Wittmeyer for sample preparation and data collection for many of our analyses; N. Blair, S. Brumfield, N. Equall, B. Glaspey, L. Kellerman, J. Monds, R. Mecham, and M. Tientze; M. Franklin, C. Paden, R. Wilkinson, and J. Starkey for lab facilities; M. Dykstra, J. Phillips, and W. Savage for imaging; W. Zheng for supporting data; and the Museum of the Rockies field crew responsible for the recovery of MOR 1125, "B. rex." Supported by NSF grants EAR-0541744 and EAR-0548847 and the David and Lucile Packard Foundation (M.H.S.), NASA Experimental Program to Stimulate Competitive Research grant NCC5-579 (R.A.), NSF grant EAR-0634136 (J.M.A.), and N. Myrhvold (J.R.H.).


### Supporting Online Material

www.sciencemag.org/cgi/content/full/316/5822/277/DC1
Materials and Methods
Figs. S1 to S7
References

11 December 2006; accepted 19 March 2007
10.1126/science.1138709

# Protein Sequences from Mastodon and *Tyrannosaurus Rex* Revealed by Mass Spectrometry

John M. Asara,[1,2]* Mary H. Schweitzer,[3] Lisa M. Freimark,[1] Matthew Phillips,[1] Lewis C. Cantley[1,4]

**Fossilized bones from extinct taxa harbor the potential for obtaining protein or DNA sequences that could reveal evolutionary links to extant species. We used mass spectrometry to obtain protein sequences from bones of a 160,000- to 600,000-year-old extinct mastodon (*Mammut americanum*) and a 68-million-year-old dinosaur (*Tyrannosaurus rex*). The presence of *T. rex* sequences indicates that their peptide bonds were remarkably stable. Mass spectrometry can thus be used to determine unique sequences from ancient organisms from peptide fragmentation patterns, a valuable tool to study the evolution and adaptation of ancient taxa from which genomic sequences are unlikely to be obtained.**

Obtaining genome sequences from a number of taxa has dramatically enhanced our abilities to study the evolution and adaptation of organisms. However, difficulties in the acquisition of DNA or RNA from ancient extinct taxa limit the ability to examine molecular evolution. Recent advances in mass spectrometry (MS) technologies have

made it possible to obtain sequence information from subpicomolar quantities of fragmented proteins and peptides (*1*, *2*), but the conversion of these fragmentation patterns (MS/MS spectra) into peptide sequences in the absence of genomic and protein sequences from publicly available databases has been a challenge. If the unknown peptide is identical in sequence to a protein region from an organism whose genes or proteins have previously been sequenced, then the fragmentation pattern (the mass/charge ratios and relative intensities of peaks) will match a theoretical fragmentation pattern from a sequence in publicly available protein databases or from the fragmentation pattern of a synthetically derived peptide, to confirm its identity.

Using this approach, tryptic peptide sequences of collagen have been identified from a 100,000- to 300,000-year-old mammoth skull, and these matched collagen fragments of extant mammalian taxa including bovine, a result that was also supported by immunological methods (*3*). MS has also been used to report protein sequences from younger fossil specimens (*4*–*6*). However, sequence data from very old (1 million years or older) fossils has been hindered by protein concentrations below the limits of detection by most analytical methods, and by theoretical limits based on predicted rates of degradation (*7*, *8*). In addition, most commercial software for identifying peptide sequences by MS relies on the peptide fragmentation pattern matching identically to that of a peptide/protein sequence in existing sequence databases. Here we show that these hindrances can be overcome by a two-step proteomics approach to obtain sequences from ion-trap MS fragmentation patterns.

We sequenced collagen protein fragments derived from fossilized bones of two extinct taxa: a 160,000- to 600,000-year-old mastodon [specimen number Museum of the Rockies (MOR) 605] (*9*) and a 68-million-year-old dinosaur (*Tyrannosaurus rex*, MOR 1125) (*10*), results that are supported by immunological and molecular analyses published in this issue by Schweitzer *et al.* (*11*). We first looked for tryptic peptide fragments from extracts of fossilized bone that matched identically with sequences from an orthologous protein or proteins from extant taxa, thereby identifying the protein(s) of interest. This is a common procedure for conserved proteins from taxa that share genomic information. Next, we generated a protein sequence database of likely drifts in amino acids in other tryptic peptides by comparing amino acid sequences of the orthologs

[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; North Carolina Museum of Natural Sciences, Raleigh, NC 27601, USA; and Museum of the Rockies, Montana State University, Bozeman, MT 59717, USA. [3]Department of Marine, Earth and Atmospheric Sciences, North Carolina State University, Raleigh, NC 27695, USA. [4]Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA.

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

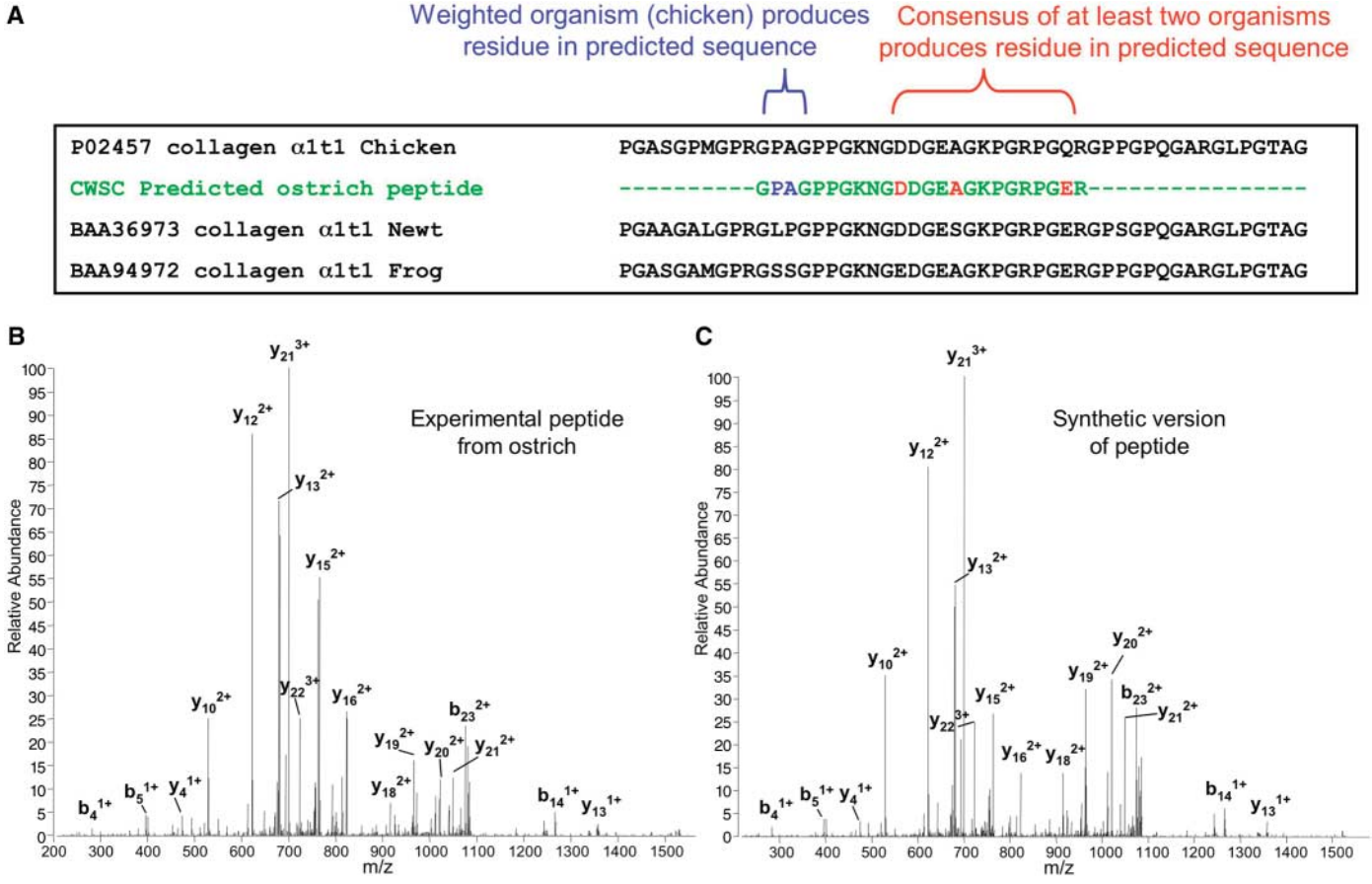


**Fig. 1.** Sequence identification by matching a peptide fragmentation pattern to a predicted tryptic peptide sequence. (**A**) Example of the chicken weighted simple consensus (CWSC) sequence algorithm for predicting a tryptic collagen peptide sequence from a previously unsequenced taxon (ostrich) based on three related organisms with one weighted organism (chicken). If a consensus of at least two organisms was present at amino acid residues that diverged (positions 11, 15, and 23), the consensus residue was chosen in the predicted tryptic peptide sequence. For residues that diverged where no consensus was present (positions 2 and 3), the residue from the weighted organism (chicken) was chosen for the predicted sequence. Amino acid residues that aligned through all three organisms were left unchanged in the predicted sequence. (**B**) The experimental MS/MS spectrum from the LC/MS/MS analysis of a triply charged peptide from ostrich bone protein extract that matched to the predicted sequence GPAGP(OH)PGKNGDDGEAGKP(OH)GRP(OH)GER and contained three hydroxyproline residues. (**C**) The MS/MS spectrum of the synthetically derived triply charged peptide of the same sequence for confirmation. The typical b- and y- fragment ions from the fragmentation pattern of the experimental peptide align very well with the synthetic peptide, validating the sequence interpretation.

from multiple related extant taxa. This approach produced a manageable number of theoretical protein sequences. The predicted peptide fragmentation pattern from these theoretical protein sequences were then compared with the fragmentation patterns of additional peptides derived from extracts of fossilized bone that did not match peptides in public sequence databases (fig. S1).

As a proof of concept for this approach, we investigated collagen sequences from femur bone extracts of an ostrich (*Struthio camelus*), an extant organism whose genomic sequence has not yet been evaluated and whose protein sequences are not available in protein databases. Collagens are the most abundant proteins in bone (>90%) and have specific posttranslational modifications (*12*, *13*), and their longevity has already been demonstrated in fossils (*3*, *14*). Collagen proteins are also highly conserved. For example, the sequence identity for collagen α1 type 1 (α1t1) from human (*Homo sapiens*) to frog (*Xenopus laevis*) is 81%, and the sequence identity between human and bovine (*Bos taurus*) is 97%, an extraordinarily high similarity. Using the Sequest algorithm (*15*), we identified 87 tryptic peptide spectra by microcapillary LC/MS/MS (LC, liquid chromatography) representing peptide sequence matches to extant related organisms in protein databases, primarily collagen α1t1 from chicken (*Gallus gallus*) (table S1).

**Table 1.** Extinct mastodon collagen proteins identified by LC/MS/MS, showing identity to extant organisms. A list of collagen proteins from 160,000- to 600,000-year-old mastodon fossilized bone is shown, including the number of peptide spectra identified, the amino acid coverage, and the organism identity.

| Protein name | Organism identity | Number of peptide spectra | Amino acid coverage |
|---|---|---|---|
| Collagen α1(I)-chain precursor | Dog, bovine, human, chimp | 24 | 20% |
| Similar to α2t1 collagen | Dog, human | 15 | 10% |
| Similar to α2t1 collagen | Elephant | 12 | 9% |
| Similar to collagen α1(IV)-chain precursor | Bovine | 3 | 4% |
| α1t1 collagen | Human | 2 | 2% |
| α1t2 collagen isoform 1 | Human | 3 | 4% |
| Collagen α2(I) chain | Human | 4 | 6% |
| Similar to collagen α1t1 | Elephant | 2 | 3% |
| Collagen α1(I) chain | Mouse | 2 | 2% |
| α1t2 collagen | Newt | 2 | 5% |
| Similar to α2t1 collagen | Chicken | 3 | 4% |
| Similar to collagen α1(I)-chain precursor | Chimp | 2 | 2% |
| α1t1 collagen | Newt | 2 | 3% |

**A**

| Peptide Sequences Unique to Mastodon | Protein | Xcorr | Sp |
|---|---|---|---|
| GSEGPQGTR | collagen α1t1 | 2.22 | 1187 |
| GAPGPQGP*G*GAP*GPK | collagen α1t1 | 3.44 | 1017 |
| EGAPGSEGAPG*RDGAIGPK | collagen α1t1 | 2.86 | 580 |
| GLTGPIGPP*GPAGAP*GDKGEG*GPSGPAGPTGAR | collagen α1t1 | 5.63 | 715 |

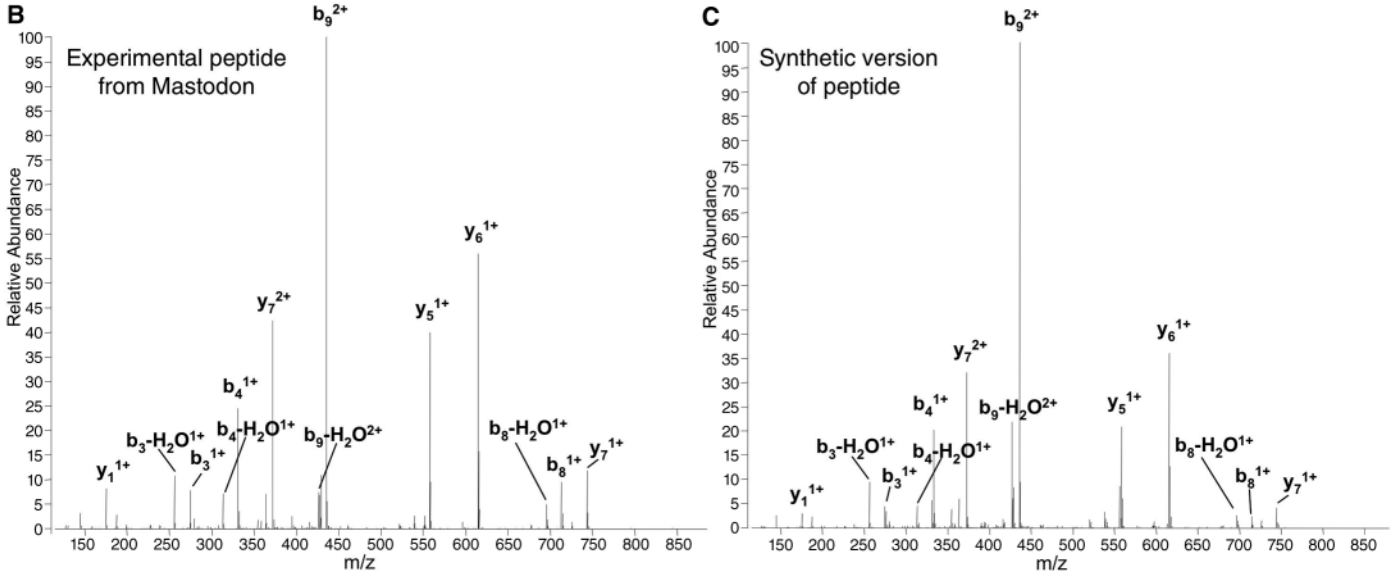


**Fig. 2.** Collagen peptide sequences unique to extinct mastodon identified by LC/MS/MS. (**A**) The four collagen α1t1 peptide sequences found by the approach that are unique to ancient mastodon. Xcorr (cross-correlation score) and Sp (preliminary score) represent the scores resulting from database searching against protein databases using Sequest. The asterisk represents the hydroxylation site after the posttranslationally modified residue. (**B**) An example of the experimental MS/MS spectrum of a doubly charged tryptic peptide for the collagen α1t1 peptide sequence GSEGPQGTR from the LC/MS/MS analysis of mastodon fossilized bone extract identified from a Sequest search against a theoretical collagen protein database. (**C**) The synthetic version of the same peptide sequence. All major ions from the experimental spectrum align very well with the ions from the synthetic version, validating the sequence.

In addition to obtaining sequence data, a further benefit of using MS is that posttranslational modifications (*16*) of the proteins can be determined. Approximately 50% of proline residues, 15% of lysines, and 10% of glycines in the collagen peptides were hydroxylated. Hydroxyproline stabilizes the triple helical confirmation of collagen in fibrillar structures (*12*), and hydroxylysine cross-links individual collagen molecules (*13*), although the function of glycine hydroxylation has not been reported. In some cases, the technology used here could not determine posttranslational modifications resulting in very small mass shifts nor could it distinguish isobaric amino acid residues. Approximately 33% of the sequence for collagen α1t1 and 16% for collagen α2t1 were identified from ostrich bone extracts through identical matches to collagen in taxa whose sequences are present in protein databases. Many experimental factors can result in a failure to obtain complete protein coverage, including proteolysis, chromatography, and ionization efficiency; however, some peptide sequences could have been missed because of the evolutionary divergence of amino acids from sequences of taxa in current protein databases.

To address this latter possibility, we generated aligned sequences obtained from chicken collagen α1t1 (the most closely related sequence to that of ostrich in the public database) with those from frog and Japanese newt (*Cynops pyrrhogaster*) (the next most closely related sequences in the database). For predicted tryptic fragments where one or more of these three taxa diverged at more than one residue, we generated a set of theoretical peptide/protein sequences that included the exact sequence in regions where all three species were identical and various combinations of the observed variant amino acids at residues where the three species diverged (Fig. 1A). We assumed that differences between chicken and ostrich are most likely to occur at residues that have been observed to drift from chicken to frog and newt. Because chicken is phylogenetically closer to ostrich than frog or newt, we chose the residue observed in chicken as the most likely residue when all three species differed at a location but chose the majority residue where two out of three were identical at a given position. For this example, we predicted a theoretical drift from chicken that was likely to be observed in a related species such as ostrich. This sequence (as well as other sequences predicted in an analogous manner) matched MS/MS fragmentation patterns from ostrich bone extract as well as a synthetic peptide created for sequence validation (Fig. 1, B and C).

Additional theoretical sequences were generated for misaligned residues by using point-assisted mutation (PAM) matrices that predict changes in amino acid residues through evolution (*17*), rather than choosing a residue from one of the three initial organisms (*18*). In addition to sequence matches to related organisms, the approach found six additional collagen peptide sequences that were unique to extant ostrich

**Table 2.** 68-million-year-old *T. rex* collagen peptide sequences identified by LC/MS/MS. Organism identity indicates the extant organisms to which the MS/MS fragmentation pattern perfectly aligned. Xcorr, Sequest cross-correlation score; Sp, Sequest preliminary score; *, hydroxylation site after a modified residue. The majority of collagen sequence matches from *T. rex* align uniquely with chicken from publicly available protein databases.

| Peptide sequence | Protein | Organism identity | Xcorr | Sp |
|---|---|---|---|---|
| GATGAP*GIAGAPG*FP*GAR | Collagen α1t1 | Chicken, frog | 3.77 | 1099 |
| G*AAGPP*GATGFP*GAAGR | Collagen α1t1 | Newt, fish, mouse | 3.74 | 797 |
| GVQGPP*GPQGPR | Collagen α1t1 | Chicken | 2.54 | 865 |
| GLPGESGAVGPAGPIGSR | Collagen α2t1 | Chicken | 2.99 | 479 |
| GVVGLP*GQR | Collagen α1t1 | Multiple organisms | 2.55 | 500 |
| GLVGAPGLRGLPGK | Collagen α1t2 | Frog | 2.28 | 410 |
| GAPGPQG*PAGAP*GPK | Collagen α1t1 | Newt | 2.14 | 272 |

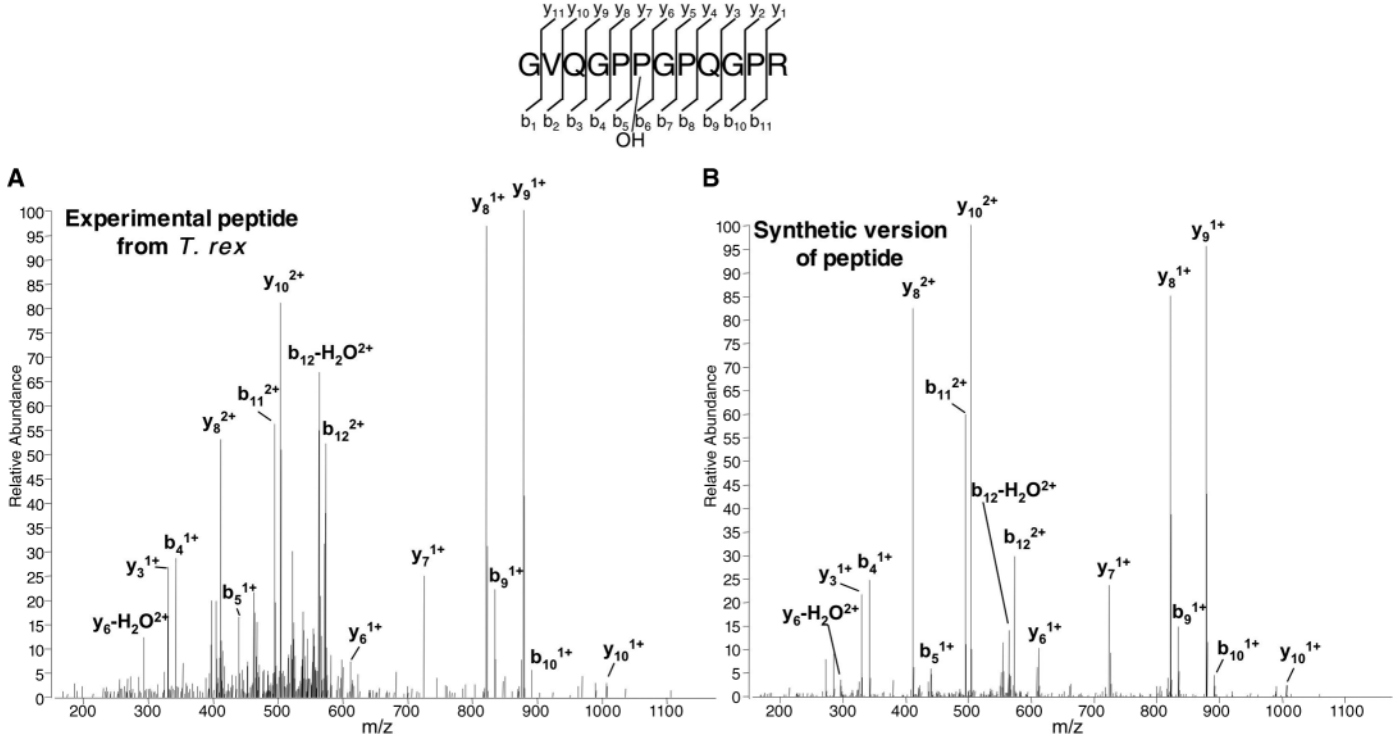


**Fig. 3.** The LC/MS/MS fragmentation pattern from a 68-million-year-old *T. rex* peptide. (**A**) The experimental MS/MS spectrum for the *T. rex* doubly charged hydroxylated tryptic peptide sequence GVQPP(OH)GPQGPR from femur bone extract identified by LC/MS/MS. (**B**) The synthetic version of the same sequence. All major fragment ions from the experimental spectrum are in very good alignment with ions from the synthetic version, confirming the sequence. This molecular sequencing evidence of protein from a 68-million-year-old fossilized bone demonstrates excellent preservation of the *T. rex* femur and the high sensitivity of state-of-the-art MS technology.

and had been missed in the comparison with public databases (table S2). Of these six sequences, four were determined using the organism weighted-consensus method and two were determined using the PAM weighted-consensus method. Although this approach only increased the total sequence coverage of ostrich collagen from ~33% to ~39%, it revealed sequences in ostrich collagen that differed from those in chicken or in other species in the database, thus providing a means to assess evolutionary divergences between these species.

For mastodon sequences, we sampled extracts from mastodon long-bone fragments dating to 160,000 to 600,000 years ago that preserved soft tissues (*9*, *19*). 74 tryptic peptide MS/MS spectra matched collagen sequences from extant mammalian organisms in the protein database (Table 1). The sequence matches resulted in approximately 32% coverage for collagen α1t1 and 29% for collagen α2t1. The fraction of glycine with hydroxylation in the mastodon collagen peptides was somewhat higher (~50%) than that observed in ostrich collagen, whereas the fraction of hydroxylated proline and lysine residues was similar to that in ostrich. It is possible that the subset of peptides with hydroxylation on glycine is resistant to proteolysis, thus explaining the enrichment. Alternatively, enzymatic glycine hydroxylation may be more active in the mastodon than in the ostrich, or the oxidation may have occurred nonenzymatically over hundreds of thousands of years. The mastodon sequences obtained were more closely related to collagen sequences from dog, bovine, human, and elephant than to nonmammalian taxa, as expected.

We used the same approach as for ostrich to identify collagen peptide sequences unique to mastodon. We compared drifts in sequences between vertebrate mammals, including human, dog (*Canis familiaris*), and mouse (*Mus musculus*) as well as PAM to generate additional theoretical collagen sequences centered on extant mammalian species and added them to our protein database containing theoretical ostrich sequences. Elephant (*Loxodonta africana*), the most closely related extant taxon (*20*), was not used to generate predicted collagen sequences because incomplete peptide sequence fragments translated from genomic data are available in public protein databases. Other combinations of mammalian organisms such as bovine were also used to generate additional predictions. Four additional mastodon peptide sequences from fragmentation patterns matched the theoretical sequences; two from the organism weighted-consensus method and two from the PAM weighted-consensus method (Fig. 2A). Those sequences were confirmed by comparing their fragmentation patterns to those obtained with synthetic peptides (Fig. 2, B and C).

The additional sequences increased the amino acid coverage of mastodon collagen α1t1 from approximately 32% to 37%, a comparable increase to that for the ostrich collagen. We thus obtained nearly as much coverage of collagen sequence from the 160,000- to 600,000-year-old fossilized bone (37%) as from a freshly collected ostrich bone (39%).

For *T. rex* sequences, we analyzed proteins from the femur of a 68-million-year-old *T. rex* (MOR 1125) recovered from the base of the Hell Creek Formation. This bone preserves soft tissues (*10*). Previous attempts to obtain protein or DNA sequences from such ancient fossils have failed because of extremely low concentrations of organic material remaining after extraction and because of degradation or modification of the remaining organic materials (*7*). Protein extracts from the *T. rex* were prepared as for ostrich and mastodon; however, the tryptic peptides required multiple purification steps (solid-phase extraction, strong cation exchange, and reversed-phase microchromatography) in order to eliminate a rust-colored coextracting contaminant and to increase the concentration of peptidic material. Three sequences from initial LC/MS/MS experiments from the *T. rex* samples indicated the presence of the iron-containing metalloenzyme nitrile hydratase beta, derived from soil bacteria *Rhodococcus sp.* and involved in biodegradation (*21*). Two peptides from a collagen adhesion protein from *Solibacter usitatis* were also sequenced. Microbial contamination was not significant, most likely because of the deep burial of the fossilized bones in the strata of the Hell Creek Formation.

The MS/MS spectra obtained from processed *T. rex* bone extracts revealed seven total collagen peptide sequences that could be aligned with predicted fragmentation patterns of collagen α1t1, α2t1, or α1t2 sequences from extant vertebrate taxa in the public protein database (Table 2). These sequences could be reproduced from multiple LC/MS/MS experiments; however, different peptides were sequenced from five different sample preparations of *T. rex* protein extract over a 1.5-year period. The last two extractions yielded less sequence information than earlier extractions, probably because of degradation of the fossil over time after removal from its well-preserved native environment (*22*). As in the extant ostrich and extinct mastodon, most of the peptides contained hydroxyl modifications on proline, lysine, or glycine residues. Sediment and buffer control samples were analyzed, and no sequences from collagen were found, although bacterial peptides were also present in sediment.

A BLAST alignment and similarity search (*23*) of the five *T. rex* peptides from collagen α1t1 as a group against the all-taxa protein database showed 58% sequence identity to chicken, followed by frog (51% identity) and newt (51% identity). The small group of peptide sequence data reported here support phylogenetic hypotheses suggesting that *T. rex* is most closely related to birds among living organisms whose collagen sequence is present in protein databases (*24*–*26*). The collagen sequences from other closely related extant taxa such as alligator (*Alligator sinensis*) and crocodile (*Crocodylus acutus*) are not present in current protein databases. If all sequences were consistent with a single extant organism, it might indicate that the samples or our experiments were contaminated. However, we identified regions of sequence that align uniquely with multiple related vertebrate taxa in protein databases. The highly conserved nature of collagen proteins results in very limited regions that do not overlap, and the sequence alignments vary by only one or two amino acids, even in distantly related organisms. Because these peptides are all derived from the same bone matrix, one would have to make the argument for multiple contamination events from organisms, such as newt, that are not native to Hell Creek environments and have never been inside the buildings that have housed these bone samples. For further validation of the sequence data, Fig. 3 shows one of the experimental *T. rex* sequences, GVQGPP(OH)GPQGPR (*27*), that matched to chicken collagen α1t1 and the synthetic version of the peptide. The experimental spectrum shows lower signal intensity and more chemical noise than the synthetic peptide, which is not surprising because the spectra were derived from 68-million-year-old endogenous proteins. The signal intensities of the mass spectra indicate that only low or subfemtomole levels of peptides were produced from tryptic digestions of approximately 30 mg of bone protein extract. Peptide sequences unique to *T. rex* were not found, most likely because few peptides were available for sequencing as compared to the ostrich and mastodon samples. In support of these results and data shown here and by Schweitzer *et al.* (*11*), in situ localization with avian antibodies to collagen type 1 shows the presence of collagen, which disappears after treatment with collagenase (*11*).

The ability to sequence intact peptides from a 68-million-year-old source is attributed to several factors, including the exceptional preservation of the soft tissues from the Hell Creek environment, the fresh preparation of the fossil samples without curation or preservation (*22*), and the advancements in the sensitivity of MS technology over the past decade. The fact that sequenceable collagen was very abundant in the mastodon sample, which could be approximately half a million years old, also sheds light on the fact that sequenceable protein lasts much longer than 1 million years.

As technologies become more refined and protein extraction techniques are optimized, more informative material may be recovered. This holds promise for future work on other fossil material showing similar preservation, but also demonstrates a method for obtaining protein sequences from rare or endangered extant organisms whose genomes have note been sequenced. The MS- and bioinformatics-based approach we have used can be applied not only to obtain sequences from extinct organisms, but also to obtain protein sequences from extant organisms whose genomes have not been sequenced and to discover mutations in diseased tissues such as cancers.

**References and Notes**

1. M. Mann, O. N. Jensen, *Nat. Biotechnol.* **21**, 255 (2003).
2. B. Domon, R. Aebersold, *Science* **312**, 212 (2006).
3. M. H. Schweitzer *et al.*, *J. Mol. Evol.* **55**, 696 (2002).
4. C. M. Nielsen-Marsh *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 4409 (2005).
5. C. M. Nielsen-Marsh *et al.*, *Geology* **30**, 1099 (2002).
6. P. H. Ostrom *et al.*, *Geochim. Cosmochim. Acta* **70**, 2034 (2006).
7. T. Lindahl, *Nature* **362**, 709 (1993).
8. M. J. Collins *et al.*, *Geology* **28**, 1139 (2000).
9. C. L. Hill, *Quat. Int.* **142-143**, 87 (2006).
10. M. H. Schweitzer, J. L. Wittmeyer, J. R. Horner, J. K. Toporski, *Science* **307**, 1952 (2005).
11. M. H. Schweitzer *et al.*, *Science* **316**, 277 (2007).
12. A. Bhattacharjee, M. Bansal, *Int. U. Biochem. Mol. Biol. Life* **57**, 161 (2005).
13. L. Knott, A. J. Bailey, *Bone* **22**, 181 (1998).
14. N. Tuross, M. L. Fogel, P. E. Hare, *Geochim. Cosmochim. Acta* **52**, 929 (1988).
15. J. K. Eng, A. L. McCormack, J. R. Yates III, *J. Am. Soc. Mass Spectrom.* **5**, 976 (1994).
16. M. Mann, O. N. Jensen, *Nat. Biotechnol.* **21**, 255 (2003).
17. M. O. Dayhoff *et al.*, *Atlas of Protein Sequence and Structure*, vol. 5, suppl. 3 (National Biomedical Research Foundation, Washington, DC, 1978).
18. A complete description of the predicted sequence generation is available as supporting material on *Science* Online.
19. M. H. Schweitzer, J. L. Wittmeyer, J. R. Horner, *Proc. R. Soc. London Ser. B* **274**, 183 (2007).
20. H. Yang, E. M. Golenberg, J. Shoshani, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 1190 (1996).
21. M. M. Tauber *et al.*, *Appl. Environ. Microbiol.* **66**, 1634 (2000).
22. M. Pruvost *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 739 (2007).
23. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, *J. Mol. Biol.* **215**, 403 (1990).
24. P. M. O'Connor, L. P. Claessens, *Nature* **436**, 253 (2005).
25. F. V. Mariani, G. R. Martin, *Nature* **423**, 319 (2003).
26. M. H. Schweitzer, J. L. Wittmeyer, J. R. Horner, *Science* **308**, 1456 (2005).
27. 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.



28. We thank J. Wittmeyer, J. Luo, R. Mecham, M. Berne, M. Vander Heiden, H. Christofk, and B. Zheng for experimental help and helpful discussions; W. Lane and S. Nagpal for software help; J. R. Horner for providing *T. rex* and mastodon fossils; and C. Hill for taphonomic data for mastodon. We also thank NSF (awards EAR-0634136 and EAR-0541744), the Paul F. Glenn Foundation, and the David and Lucile Packard Foundation for providing funds.


**Supporting Online Material**

www.sciencemag.org/cgi/content/full/316/5822/280/DC1
Materials and Methods
SOM Text
Fig. S1
Tables S1 and S2
Reference


15 November 2006; accepted 16 March 2007
10.1126/science.1137614


# Lymphotoxin β Receptor–Dependent Control of Lipid Homeostasis


**James C. Lo,[1]* Yugang Wang,[2]* Alexei V. Tumanov,[2]* Michelle Bamji,[3] Zemin Yao,[3] Catherine A. Reardon,[2] Godfrey S. Getz,[2]† Yang-Xin Fu[1,2]†**



Hyperlipidemia, one of the most important risk factors for coronary heart disease, is often associated with inflammation. We identified lymphotoxin (LT) and LIGHT, tumor necrosis factor cytokine family members that are primarily expressed on lymphocytes, as critical regulators of key enzymes that control lipid metabolism. Dysregulation of LIGHT expression on T cells resulted in hypertriglyceridemia and hypercholesterolemia. In low-density lipoprotein receptor–deficient mice, which lack the ability to control lipid levels in the blood, inhibition of LT and LIGHT signaling with a soluble lymphotoxin β receptor decoy protein attenuated the dyslipidemia. These results suggest that the immune system directly influences lipid metabolism and that LT modulating agents may represent a novel therapeutic route for the treatment of dyslipidemia.


Atherosclerosis results from a combination of dyslipidemia (dysregulation of lipid levels in the blood) and inflammation-mediated pathology of the vasculature (*1*–*3*). In particular, systemic inflammation is often associated with hyperlipidemia, although the exact mechanistic link is unclear (*4*). Members of the tumor necrosis factor superfamily (TNFSF) of cytokines are involved in a wide array of biological processes, including immune function, cell death and survival, and atherosclerosis. Among these, lymphotoxin (LT) and LIGHT are potent proinflammatory ligands expressed primarily on T cells, and both bind to lymphotoxin β receptor (LTβR), a tumor necrosis factor receptor superfamily member. Disruption of LTβR signaling in mice results in a wide array of phenotypes including lymph node aplasia, impaired splenic microarchitecture, autoimmunity, viral infection, and grossly impaired immunoglobulin (Ig) responses (*5*–*7*).

Recent clinical, genetic, and epidemiological studies have linked LT to the development of atherosclerosis and myocardial infarction (*8*, *9*). However, the contribution of LT to lipid homeostasis—a likely influencing factor in atherogenesis—is not established. In one inflammatory mouse model, increased expression of LIGHT on T cell lineage in *Lck* LIGHT transgenic (Tg) mice culminates in the development of systemic inflammation, hepatomegaly, and inflammatory bowel disease (*10*, *11*). Lymphocytes can have intimate contact with hepatocytes, which raises the possibility that some T cell–derived ligands such as LT and LIGHT may deliver signals that regulate various liver functions. Indeed, LT and LIGHT have been shown to impart signals necessary for liver regeneration as well as for pathogenesis of hepatitis (*12*, *13*).

Given the observed effects of LT and LIGHT on the liver, and given that lipid homeostasis

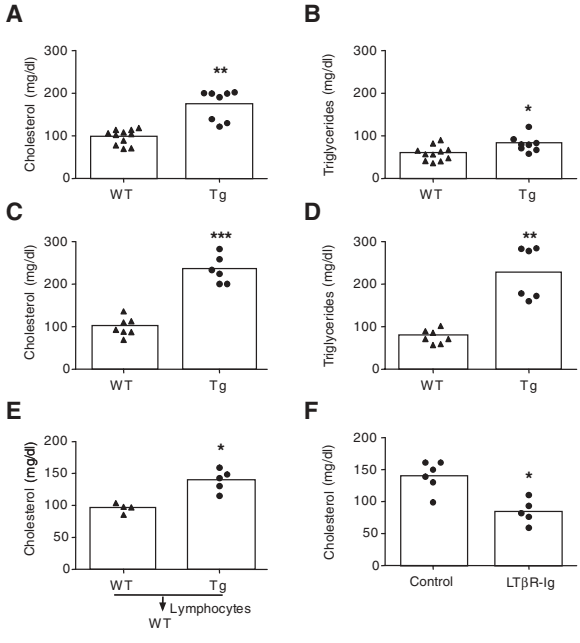


**Fig. 1.** T lymphocyte–derived LIGHT promotes dyslipidemia. (**A** to **D**) Wild-type (WT) and LIGHT Tg mice were fed a normal chow diet (**A** and **B**) or Western-type diet (WTD) (**C** and **D**) for 2 weeks and then assayed for plasma total cholesterol and triglycerides. (**E**) WT and LIGHT Tg spleen and lymph node cells ($10^7$ per mouse) were transferred into sublethally irradiated (600 rad) WT mice that were then fed WTD; after 4 weeks, plasma total cholesterol levels were determined. (**F**) LIGHT Tg mice on a normal chow diet were injected with control human Ig (hIg) or LTβR-Ig on day 0; after 1 week, plasma total cholesterol was determined. Each individual symbol represents values from WT (triangles) and LIGHT Tg (circles) mice. Means are plotted as columns. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$.


[1]Committee on Immunology, University of Chicago, Chicago, IL 60637, USA. [2]Department of Pathology, University of Chicago, Chicago, IL 60637, USA. [3]Department of Biochemistry, Microbiology and Immunology, University of Ottawa Heart Institute, Ottawa, Ontario K1H 8M5, Canada.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: yfu@uchicago.edu (Y.-X.F.); getz@bsd.uchicago.edu (G.S.G.)