Published in final edited form as:

J Proteome Res. 2009 September ; 8(9): 4328–4332. doi:10.1021/pr900349r.

Reanalysis of *Tyrannosaurus rex* Mass Spectra

Marshall Bern^{*}, Brett S. Phinney[†], and David Goldberg^{*}

Marshall Bern: bern@parc.com; Brett S. Phinney: ; David Goldberg:

^{*}Palo Alto Research Center, 3333 Coyote Hill Rd., Palo Alto, CA 94304

[†]Genome Center, University of California at Davis, Davis, CA 95616

Abstract

Asara et al. reported the detection of collagen peptides in a 68-million-year-old *T. rex* bone by shotgun proteomics. This finding has been called into question as a possible statistical artifact. We reanalyze Asara et al.'s tandem mass spectra using a different search engine and different statistical tools. Our reanalysis shows a sample containing common laboratory contaminants, soil bacteria, and bird-like hemoglobin and collagen.

Keywords

dinosaur; fossil; collagen; hemoglobin; contamination

1 Introduction

Asara et al. [1] reported finding seven distinct collagen sequences by shotgun proteomics in a remarkably well-preserved 68-million-year-old fossilized dinosaur bone. They later rejected one of the sequences as statistically insignificant, corrected the placement of hydroxylations on two other sequences [2,3], and computed a phylogenetic tree that (unsurprisingly) placed *T. rex* with birds [13]. The discovery of intact protein in such an ancient sample has been called into question on the grounds of plausibility [8] and inadequate statistical analysis [14].

In a reanalysis of the data, Matt Fitzgibbon and Martin McIntosh (supplementary material) pointed out that five of the six *T. rex* sequences appear in ostrich collagen. (Ostrich collagen has not been sequenced, but Asara et al. [13] deduced partial sequence by mass spectrometry and a computational mutation search using chicken.) The close similarity with ostrich and the fact that the same proteomics laboratory processed both *T. rex* and ostrich raised a third issue: contamination from another proteomics sample. Moreover, Fitzgibbon and McIntosh found a spectrum matching a bird hemoglobin peptide with a carbamidomethylated cysteine. This bolstered the case for exogenous contamination, because Asara et al. did not originally describe alkylation as part of their sample preparation.

The actual laboratory procedures, however, were somewhat complex. As explained by John Asara (personal communications), the entire data set comprises seven chromatographic runs of *T. rex* bone and four chromatographic runs of sediment from around the bone, chipped away during excavation. The runs were made over the course of more than a year, and some of the *T. rex* injections were alkylated to produce carbamidomethylated cysteine and some were not. The confident collagen sequences appear in several different runs and none contain cysteine. The ostrich runs were performed more than a year before any of the *T. rex* runs, and more than 1000 other runs were performed in between.

Here we reanalyze the original mass spectra yet again using different bioinformatics tools and statistical tests. We find three distinct collagen peptides matched with E-value below 1.0, along with a number of less significant matches to collagen. Assuming statistical independence of distinct peptides, the identification of bird-like collagen at the protein level is clearly significant. We also confirm the statistical significance of the bird hemoglobin peptide reported by Fitzgibbon and McIntosh.

2 Methods

We obtained the entire *T. rex* data set as an MGF (Mascot generic format) file containing 31,367 distinct MS/MS (Thermo LTQ) spectra and 48,216 combinations of spectrum and precursor charge assignment; these spectra are publicly available. Because precursor charge assignments for ion-trap spectra can be unreliable, we ignored the charge assignments in the MGF file and considered each of the spectra with assignments of +1, +2, and +3. For our protein database, we used **uniprot_sprot.fasta** (downloaded 4 February 2008), containing 408,099 protein sequences.

We searched the spectra against the database using ByOnic [5] and compiled a protein list using the companion program ComByne [6]. Previous searches used Sequest [1] and Mascot [4]. ByOnic uses a “matched filter” or dot-product scorer, incorporating predicted and observed peak intensities and mass deviations between predicted and observed peaks. In previous studies [5,6], we have found ByOnic to be more sensitive than Mascot and Sequest at the same false discovery rate. We initially performed a **wide search**, in which we searched all of **uniprot_sprot.fasta** for fully tryptic peptides, with 1500 ppm precursor mass tolerance and 0.4 Dalton fragment mass tolerance, with the following modifications enabled: carbamidomethylated cysteine (camC, a fixed modification), hydroxyproline (a common modification in collagen), oxidized methionine, and pyro-glu from N-terminal glutamine, glutamic acid, and camC. Not all of the seven chromatographic runs used camC, so we also searched the data assuming unmodified cysteine. We allowed any number of missed cleavages. We also searched a decoy database containing reversals of the protein sequences in **uniprot_sprot.fasta**. We then made a small database containing all the proteins (forward or reverse) with at least one match scoring at least 200, approximately equivalent to Mascot 20. We also added reversals of all the forward proteins in the small database, and in total the small database contained 4472 forward proteins and 7161 reversed proteins.

It may seem strange to have unequal numbers of forward and reversed proteins, but we did this to guard against bias. The assumption behind the forward/reversed database approach [9,12] is that the false positives are equally divided between the forward and reversed proteins. If our small database had included only the forward proteins from the first search along with their reversals, we would have matched the numbers and lengths of forward and reversed proteins, but our forward proteins would have been statistically different from our reversed proteins, because the forward proteins had already found matches scoring at least 200. If our small database had included only those proteins, forward or reverse, with matches of at least 200 in the first search, our small database would have been unbiased at the protein level, even though the forward proteins would have outnumbered the reverse proteins. The small database, however, would have been biased at the peptide level, likely to contain more false forward peptides than false reversed peptides, because not every peptide in a true protein is true, that is, truly the peptide represented by a mass spectrum. By including both the first-search reversed proteins and reversals of the first-search forward proteins, we slightly bias against forward peptides; this is the conservative approach.

For our **narrow search** we searched the small database for fully tryptic peptides, with the same mass tolerances and modifications as above, along with deamidated asparagine and glutamine

and at most one SNP mutation per peptide. We searched the small database one more time, using the same modifications along with a “wild-card modification” in order to ensure that nothing interesting was overlooked. By-Onic's wild-card modification allows any integer mass change to any one residue. A peptide can carry both known modifications and a wild card, and the total mass of modifications must lie within a user-settable range, in this case -50 to +80 Daltons.

Rather than relying only on ByOnic and ComByne's internal p-value computations, which were built with training data from a variety of instruments and samples, we also estimated statistical significance with an *empirical E-value* specific to this data set. For both the wide and narrow searches, we estimated the expected number of random identifications with a given ByOnic score by running a search against a database containing only decoy proteins (reversed sequences). This search used the same mass spectra and precursor charge assignments, and searched the decoy database using the same modifications and cleavage specificity.

After searching the spectra from the bone sample, we obtained from John Asara 7085 MS/MS (Thermo LTQ) spectra from a sample of the sediment surrounding the *T. rex* fossil bone. The sediment serves as a control to guard against the possibility that proteins from the environment could be mistaken for proteins from the bone fossil. We searched these spectra in the same way as our wide search, again using all protein sequences in **uniprot_sprot.fasta** along with reversed protein sequences.

3 Results and Discussion

ComByne's list of protein groups from the wide search on the bone sample, ranked by logarithm of the p-value (confidence), is shown in Table 1. This list is from the search assuming camC; without this assumption hemoglobin subunit alpha at rank 9 is lost. ComByne's log p-value reflects protein length along with the p-values for all the spectra matched to that protein, so according to ComByne's probabilistic model the matches to the top protein should appear by chance with probability $10^{-48.05}$. Assuming statistical independence of distinct collagen peptides, the matches to the top two collagens should appear by chance with probability $10^{-5.47-4.96}$ or about 10^{-10} . The assumption of independence may be questionable because collagen is highly self-similar, but even similar sequences generally have quite different mass spectra, as a single substitution shifts roughly half of the ion peaks. The best log p-value achieved by any of the 408,099 decoy proteins was -4.60 and only two other decoys had log p-value below -2.89, so we would expect about three false positives among the 40 proteins in Table 1, for a false discovery rate of about 7.5%. ComByne's p-values are generally more conservative than empirical significance testing: we would expect about 400 out of 400,000 random proteins to have p-value below 0.001 and log p-value below -3.0. The multiorganism protein database, however, is quite redundant, so its effective size is much lower than 400,000.

Table 2 gives the peptide identifications for all the collagen and hemoglobin matches from either the wide or narrow search with ByOnic scores above 250, which is roughly equivalent to a Mascot score of 25. A match with ByOnic score of 300 has an empirical E-value about 100 in both the narrow and wide searches (Figure 1), so it could be discounted as statistically insignificant, yet the probability of any given spectrum hitting a collagen protein by chance within either database, large or small, is less than 0.01, which if factored into the calculation, would bring the E-value of such a match to below 1. (Again there is an assumption of independence: the protein identity and the ByOnic score are statistically independent.) Indeed reversed collagen and hemoglobin sequences received no matches with scores above 250. The number of spectra drops roughly linearly on the logarithmic scale of Figure 1; others have observed that the right-hand tail of the score distribution for most peptide scorers is well modeled by an exponential distribution [6,10].

Asara et al. deposited six *T. rex* collagen sequences in GenBank: GATGAPGIAGAPGFPGARGAPG-PQGPGAPGPK, GSAGPPGATGFPGAAGR, GVQGGPGQGPR, and GVVGLPGQR from collagen alpha-I type I, GLVGAPGLRGLPGK from collagen alpha-1 type II, and GLPGESGAVGPAGPIGSR from collagen alpha-2 type I. Of these, we confirm the first three sequences with E-values below 1.0. We find the fourth sequence with E-value on the order of 10.0, and can argue for its correctness based on the unlikelihood of hitting collagen at random as above. We do not find the last two sequences and suggest that they be dropped from GenBank.

Several of the sequences in Table 2, for example, GLAGPQGPR, PGPQGPGAP[+16]GPK, GAPG-PQGP[+16]AGAPGP[+16]K, GVVGLP[+16]GQR, and P[+16]GC[+57]P[+16]GPMGEK, do not appear in the published partial ostrich sequence [13], but the published sequence has only about 30% coverage of collagen alpha-1, so we cannot rule out ostrich. The peptide P[+16]GC[+57]P[+16]GPMGEK is from mammalian collagen alpha-4, which is found in extracellular matrix but not bone. This hit (score 345) could be a false match, or it could be a collagen alpha-1 or alpha-2 peptide from an unsequenced organism, as all the collagens are similar at the sequence level.

As reported by Asara et al., the proteins in Table 1 are mainly known contaminants in biochemistry laboratories (human keratin, bovine serum albumin, etc.), proteins from soil bacteria (*Acidovorax*, *Verminephrobacter*, *Polaromonas*, *Zymomonas*, etc.) and other organisms plausibly found in soil, such as *Schizosaccharomyces* (a yeast), *Physcomitrella* (a moss), and *Neospora caninum* (a parasite infecting dogs and cattle). *Arachis hypogaea* (peanut) allergen appears out of place, but the matches to peanut may also match some protein in *Hevea brasiliensis*, the source of the latex in laboratory gloves and of protein 33 in Table 1, as cross-reactivity between natural latex and peanut has been reported [11]. Also as reported by Asara et al., there are very few vertebrate proteins in Table 1, so the matches to bird-like collagen and bird hemoglobin stand out. Tubulin is a vertebrate protein, but the tubulin peptides found in the sample are also found in many invertebrates. The only high-scoring match to actin, the peptide AGFAGDDAPR, matches 273 actins in **uniprot_sprot.fasta**, indeed almost all sequenced eukaryotes.

The search of the spectra from the sediment sample (Table 3) yielded many fewer proteins, with keratins and trypsin the only statistically significant finds in this relatively small data set. There were four low-scoring, statistically insignificant, hits to interesting proteins: two to vertebrate collagens, one to bird mimecan, and one to bird hemoglobin subunit beta (Q[-17] LISGLWGK from ostrich hemoglobin subunit beta with score 247). Because collagen and hemoglobin are found in the surrounding sediment in such trace quantities, if they are found at all, we do not think these hits indicate a source of collagen and hemoglobin other than the fossil bone.

In summary, we find nothing obviously wrong with the *Tyrannosaurus rex* mass spectra: the identified peptides seem consistent with a sample containing old, quite possibly very ancient, bird-like bone, contaminated with only fairly explicable proteins. Hemoglobin and collagen are plausible proteins to find in fossil bone, because they are two of the most abundant proteins in bone and bone marrow. Schweitzer et al. [15] previously reported multiple lines of evidence, including immunological reactions, for hemoglobin-derived compounds in *T. rex* bone, and collagen from younger fossil bones is well known [17]. Contamination remains a tricky and possibly unresolvable issue for this particular sample. Perhaps a bird died on top of the *T. rex* excavation in the field; perhaps ostrich bone lingered in the mass spectrometry facility for a year; or perhaps avian collagen from a cosmetic or medical product found its way into the *T. rex* sample. Complete sequencing of ostrich collagen would help dispel one contamination scenario. In just-published work on an 80-million-year-old hadrosaur fossil, Schweitzer et al.

[16] took extra precautions against contamination, including excavation with sterilized tools and analysis of the fossil extracts by more than one mass spectrometry laboratory. So far this new study has met with much less skepticism. It is fair to say that the scientific community is still working out standards for sample handling and data analysis of fossil protein.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Marshall Bern was supported in part by NIH grant GM085718. We would like to thank Pavel Pevzner for asking for the release of this intriguing data set, and John Asara for making the data available to the scientific community.

References

1. Asara JM, Schweitzer MH, Freemark LM, Phillips M, Cantley LC. Protein sequences from mastodon and *Tyrannosaurus rex* revealed by mass spectrometry. *Science* April 13;2007 316:280–284. [PubMed: 17431180]
2. Asara JM, Garavelli JS, Slatter DA, Schweitzer MH, Freemark LM, Phillips M, Cantley LC. Interpreting sequences from mastodon and *T. rex*. *Science* September 7;2007 317:1324–1325. [PubMed: 17823333]
3. Asara JM, Schweitzer MH. Response to comment on “Protein sequences from mastodon and *Tyrannosaurus rex* revealed by mass spectrometry”. *Science* January 4;2008 319:33d.
4. Asara JM, Schweitzer MH, Cantley LC, Cottrell JS. Response to comment on “Protein sequences from mastodon and *Tyrannosaurus rex* revealed by mass spectrometry”. *Science* August 22;2008 321:1040c.
5. Bern M, Cai Y, Goldberg D. Lookup peaks: a hybrid of de novo sequencing and database search for protein identification by tandem mass spectrometry. *Anal Chem* 2007;79:1393–1400. [PubMed: 17243770]
6. Bern M, Goldberg D. Improved ranking functions for protein and modification-site identifications. *J Comp Biology* 2008;15:705–719.
7. Bradshaw RA, Burlingame AL, Carr S, et al. Reporting protein identification data: the next generation of guidelines. *Mol Cell Proteomics* 2006;5:787–788. [PubMed: 16670253]
8. Buckley M, et al. Comment on protein sequences from mastodon and *Tyrannosaurus rex* revealed by mass spectrometry. *Science* January 4;2008 319:33. [PubMed: 18174420]
9. Elias JE, Gygi SP. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nature Methods* 2007;4:207–214. [PubMed: 17327847]
10. Fenyö D, Phinney BS, Beavis RC. Determining the overall merit of protein identification data sets: rho-diagrams and rho-scores. *J Proteome Research* 2007;6:1997–2004. [PubMed: 17397212]
11. Hovanec-Burns D, Ordonez M, Corrao M, Enjamuri S, Unver E. Identification of another latex-cross-reactive food allergen: peanut (Abstract). *J Allergy Clin Immunol* 1995;95:40.
12. Moore RE, Young MK, Lee TD. Qscore: an algorithm for evaluating SEQUEST database search results. *J Am Soc Mass Spectrom* 2002;13:378–386. [PubMed: 11951976]
13. Organ CL, Schweitzer MH, Zhang W, Freemark LM, Cantley LC, Asara JM. Molecular phylogenetics of mastodon and *T rex*. *Science* April 25;2008 320:499. [PubMed: 18436782]
14. Pevzner PA, Kim S, Ng J. Comment on protein sequences from mastodon and *Tyrannosaurus rex* revealed by mass spectrometry. *Science* August 22;2008 321:1040b. [PubMed: 18719266]
15. Schweitzer MH, Marshall M, Carron K, Bohle DS, Busse SC, Arnold EV, Barnard D, Horner JR, Starkey JR. Heme compounds in dinosaur trabecular bone. *Proc National Academy Sciences USA* June 10;1997 94:6291–6296.
16. Schweitzer MH, Zheng W, Organ CL, Avci R, Suo Z, Freemark LM, Lebleu VS, Duncan MB, Vander Heiden MG, Neveu JM, Lane WS, Cottrell JS, Horner JR, Cantley LC, Kalluri R, Asara JM.

Biomolecular characterization and protein sequences of the Campanian hadrosaur, *B. canadensis* Science May 1;2009 324:626–631.

17. Semal P, Orban R. Collagen extraction from recent and fossil bones: quantitative and qualitative aspects. J Archaeological Science 1995;22:463–467.

Figure 1.

This plot gives a histogram showing the number of decoy matches for the wide and narrow searches as a function of ByOnic score. The top hits to distinct collagen peptides have scores 509, 526, and 536, better than any decoy matches, with empirical E-values below 1. The top hit to bird hemoglobin beta has score 782, with an extrapolated E-value below 0.001.

The top 40 protein groups found by our wide search, showing protein log p-value, along with numbers of spectra and unique peptides with ByOnic score above 180. Not far off this list is rank 49, hemoglobin subunit alpha-A, matching 4 bird species, with log p-value -2.30 and 2 spectra and 2 unique peptides. ComByne ranks proteins by total log p-value, and “one-hit wonders” often rank higher than proteins with several matches.

Table 1

Rank	Protein	Organisms	Log p-val	# Spec	# Peps
1	Keratin, type I cytoskeletal 1	<i>Homo sapiens, Pan troglodytes</i> (Chimpanzee)	-48.05	16	10
2	Keratin, type II cytoskeletal 1	<i>Homo sapiens</i>	-47.09	13	9
3	Keratin, type II cytoskeletal 2 epidermal	<i>Homo sapiens</i>	-19.02	7	6
4	Serum albumin	<i>Bos taurus</i> (Cow)	-14.39	8	6
5	Beta-casein	<i>Bos taurus</i>	-13.92	3	3
6	Keratin, type II cytoskeletal 5	<i>Homo sapiens</i>	-12.81	4	4
7	ATP synthase subunit beta	<i>Acidovorax</i>	-12.30	3	3
8	ATP synthase gamma chain	<i>Polaromonas</i>	-11.10	1	1
9	Hemoglobin subunit beta	25 bird species	-11.08	1	1
10	Beta-galactosidase	<i>E. coli K12</i>	-10.59	1	1
11	Keratin, type I cytoskeletal 9	<i>Homo sapiens</i>	-10.37	8	5
12	Allergen Ara h 1	<i>Arachis hypogaea</i> (Peanut)	-9.49	3	3
13	Elongation factor Ts	<i>Bordetella, Burkholderia</i> (21 species)	-9.20	1	1
14	Keratin, type I cytoskeletal 16	<i>Homo sapiens</i>	-9.09	5	5
15	50S ribosomal protein	<i>Verminephrobacter</i>	-8.90	1	1
16	Malate dehydrogenase	<i>Citrobacter, Escherichia</i> (44 species)	-7.90	1	1
17	Dermcidin	<i>Homo sapiens</i>	-7.86	1	1
18	Beta-lactoglobulin	<i>Bos taurus, Bubalus bubalis</i>	-7.76	1	1
19	Glyceraldehyde-3-phosphate dehydr'ase	<i>Physcomitrella</i>	-7.60	1	1
20	ATP synthase subunit alpha 1	<i>Acidovorax</i> (27 species)	-7.20	1	1
21	ATP synthase gamma chain	<i>Cytophaga hutchinsonii</i>	-6.07	1	1
22	Trypsin	<i>Sus scrofa</i> (Pig)	-5.69	2	2
23	Collagen alpha-1(I)	<i>Cynops pyrrhogaster</i> (Newt)	-5.47	5	3
24	Collagen alpha-1(I)	<i>Gallus gallus</i> (Chicken)	-4.96	2	2
25	Elongation factor G	<i>Zymomonas mobilis</i>	-4.91	2	2
26	Tubulin alpha-1A chain	<i>Homo sapiens, Gallus gallus</i> , (22 species)	-4.50	2	2
27	Ig kappa chain V region	<i>Oryctolagus cuniculus</i> (rabbit)	-4.46	1	1

Rank	Protein	Organisms	Log p-val	# Spec	# Peps
28	Actin	(273 species)	-4.45	1	1
29	Acetylglutamate kinase	<i>Methylobium petroleiphilum</i>	-4.24	1	1
30	Keratin, type II cytoskeletal 6A or 6C	<i>Homo sapiens</i>	-3.84	3	3
31	Collagen alpha-1(I) chain (Fragments)	<i>Tyrannosaurus rex</i>	-3.76	1	1
32	Uncharacterized protein C57A7.05	<i>Schizosaccharomyces pombe</i>	-3.73	1	1
33	Rubber elongation factor protein	<i>Hevea brasiliensis</i> (Rubber)	-3.65	1	1
34	Dense granule protein 2	<i>Neospora caninum</i>	-3.60	1	1
35	Phosphate import ATP-binding protein	<i>Methanospirillum, Rhodospirillum</i>	-3.20	1	1
36	Heat shock cognate 70 kDa protein	<i>Bos taurus, Homo sapiens</i> (19 species)	-3.65	2	2
37	Elongation factor Tu	<i>Acidovorax, Rhodoferrax</i> (10 species)	-3.00	1	1
38	Pyruvate kinase isozyme	<i>Mus musculus, Homo sapiens</i> (13 species)	-3.00	1	1
39	Betaine aldehyde dehydrogenase	<i>Burkholderia</i>	-2.98	1	1
40	Collagen alpha-2(I)	<i>Gallus gallus</i>	-2.89	2	2

Table 2

The spectra matching collagen and hemoglobin. Asterisks show mutated sequences. The first * sequence disagrees with the higher-scoring unmutated sequences above it, so the more likely interpretation of this spectrum is K.GATGAP[+16]GIAGAP[+16]GFP[+16]GAR.G. The sequence marked † has score below 250, but we included it anyway because it is corroborated by the match above it. Of the matches shown here, this match was the only one found only by the wild-card (blind modification) search; +57 on the N-terminus is probably over-alkylation.

Scan	Peptide	Organisms	Score
0130Tmsdinoedtac.2295.2295.2.dta	R.GLAGPQGPR.G	Collagen alpha-2(I) (fragments), <i>G. Gallus</i>	258
0130Tmsdinoedtac.2357.2357.2.dta	R.GAPGPQGP[+16]AGAPGP[+16]K.G	Collagen alpha-1(I), <i>C. pyrrhogaster</i> , <i>T. rex</i>	334
0607Tmscorh2x.997.997.2.dta	R.GAPGPQGPSGAP[+16]GPK.X	Collagen alpha-1(I), <i>T. rex</i>	526
0130Tmsdinoedtac.2736.2736.2.dta	R.GPP[+16]GESGAAGPTGPIGSR.G	Collagen alpha-2(I), <i>H. sapiens</i> , <i>B. taurus</i>	257
0130Tmsdinoedtac.2784.2784.2.dta	R.GSAGPP[+16]GATGFP[+16]GAAGR.V	Collagen alpha-1(I), <i>G. Gallus</i>	509
0421Tjadinocortzipr.3000.3000.2.dta	R.GSAGPP[+16]GATGFP[+16]GAAGR.V	Collagen alpha-1(I), <i>G. Gallus</i>	404
0421Tjadinomedzip.3021.3021.2.dta	R.GSAGPP[+16]GATGFP[+16]GAAGR.V	Collagen alpha-1(I), <i>G. Gallus</i>	435
0628Tmsmor1125v.3341.3341.2.dta	R.GSAGPP[+16]GATGFP[+16]GAAGR.V	Collagen alpha-1(I), <i>G. Gallus</i>	387
0130Tmsdinoedtac.3160.3160.2.dta	K.GATGAP[+16]GIAGAP[+16]GFP[+16]GAR.G	Collagen alpha-1(I), <i>G. Gallus</i>	309
0419Tjatresxscxc18zip.3552.3552.2.dta	K.GATGAP[+16]GIAGAP[+16]GFP[+16]GAR.G	Collagen alpha-1(I), <i>G. Gallus</i>	435
0421Tjadinomedzip.3451.3451.2.dta	K.GATGAP[+16]GIAGAP[+16]GFP[+16]GAR.G	Collagen alpha-1(I), <i>G. Gallus</i>	536
0628Tmsmor1125v.3951.3951.2.dta	K.GATGAP[+16]GIAGAP[+16]GFPGA[+16]R.G	Collagen alpha-1(I), <i>G. Gallus</i> *	327
0130Tmsdinoedtac.2928.2928.2.dta	R.GL[-16]P[+16]GESGAVGPAGPIGSR.G	Collagen alpha-2(I) (fragments), <i>G. Gallus</i> *	265
0421Tjadinocortzipr.2931.2931.2.dta	R.GVVGLP[+16]GQR.G	Collagen alpha-1(I), <i>G. gallus</i> , <i>H. sapiens</i>	361
0421Tjadinocortzipr.3057.3057.2.dta	R.GEP[+16]GPAGLP[+16]GPAGER.G	Collagen alpha-1(I), <i>G. Gallus</i>	324
0607Tmscorh2x.3432.3432.2.dta	K.GATGAP[+16]GIAGAP[+16]GFP[+16]GAR.G	Collagen alpha-1(I), <i>G. Gallus</i>	389
0607Tmscorh2x.3435.3435.2.dta	K.GATGAP[+16]GIAGAP[+16]GFP[+16]GAR.G	Collagen alpha-1(I), <i>G. Gallus</i>	447
0628Tmsmor1125v.2855.2855.2.dta	R.GVQGPP[+16]GPQGPR.G	Collagen alpha-1(I), <i>G. Gallus</i>	410
0628Tmsmor1125v.2963.2963.2.dta	R.GP[+16]PGSS[-16]GSTGK.D	Collagen alpha-1(I), <i>C. pyrrhogaster</i> *	250
0628Tmsmor1125v.4350.4350.2.dta	K.GAAGLPGV[+16]AGAP[+16]GLPGP[+16]R.G	Collagen alpha-2(I) (fragments), <i>G. Gallus</i> *	292
112905Tmsd5.1489.1489.2.dta	R.P[+16]GC[+57]P[+16]GPMGEK.G	Collagen alpha-4(IV) chain, <i>M. musculus</i>	345
0130Tmsdinoedtac.3382.3382.2.dta	K.VNVADC[+57]GAEALAR.L	Hemoglobin beta A, Various birds	782
0130Tmsdinoedtac.3493.3493.2.dta	K.V[+57]NVADC[+57]GAEALAR.L	Hemoglobin, beta A, Various birds†	207
0130Tmsdinoedtac.768.768.2.dta	K.LSDLHAQK.L	Hemoglobin, alpha A, Various birds	398

The top 12 protein groups found by wide search on 7085 spectra from a sample of sediment collected along with the *T. rex* bone. All but the first 5 are statistically insignificant; protein 6 ranks below the top two reversed sequences.

Table 3

Rank	Protein	Organisms	Log p-val	# Spec	# Peps
1	Keratin, type I cytoskeletal 10	<i>Homo sapiens</i>	-34.20	12	11
2	Keratin, type II cytoskeletal 1	<i>Homo sapiens</i> , <i>Pan troglodytes</i>	-20.54	11	9
3	Keratin, type I cytoskeletal 9	<i>Homo sapiens</i>	-13.18	4	4
4	Keratin, type II cytoskeletal 2 epidermal	<i>Homo Sapiens</i>	-7.88	4	4
5	Trypsin	<i>Sus scrofa</i> (Pig)	-5.86	5	2
6	Leucine-rich repeat-containing protein 42	<i>Rattus norvegicus</i> (Rat)	-2.04	2	2
7	Ezrin-radixin-moesin-binding phosphoprotein 50	<i>Homo sapiens</i>	- 1.91	2	1
8	Serine/arginine repetitive matrix protein 2	<i>Mus musculus</i>	-1.50	2	1
9	Uncharacterized protein C5E4.10c	<i>Schizosaccharomyces pombe</i>	-1.50	1	1
10	Bifunctional protein glmU	<i>Pseudomonas putida</i>	-1.50	1	1
11	Keratin, type II cytoskeletal	<i>Homo sapiens</i>	-1.50	1	1
12	V-type proton ATPase catalytic subunit A	<i>Neurospora crassa</i> (Mold)	-1.48	1	1