Proteomics in the Cretaceous

John M. Asara, Ph.D.

Mass Spectrometry Core
Beth Israel Deaconess Medical Center
Harvard Medical School
Boston, MA

The choice is always tempting in our mass spec facility…..

Multi-million year old proteins from fossil bones

15-20% effort

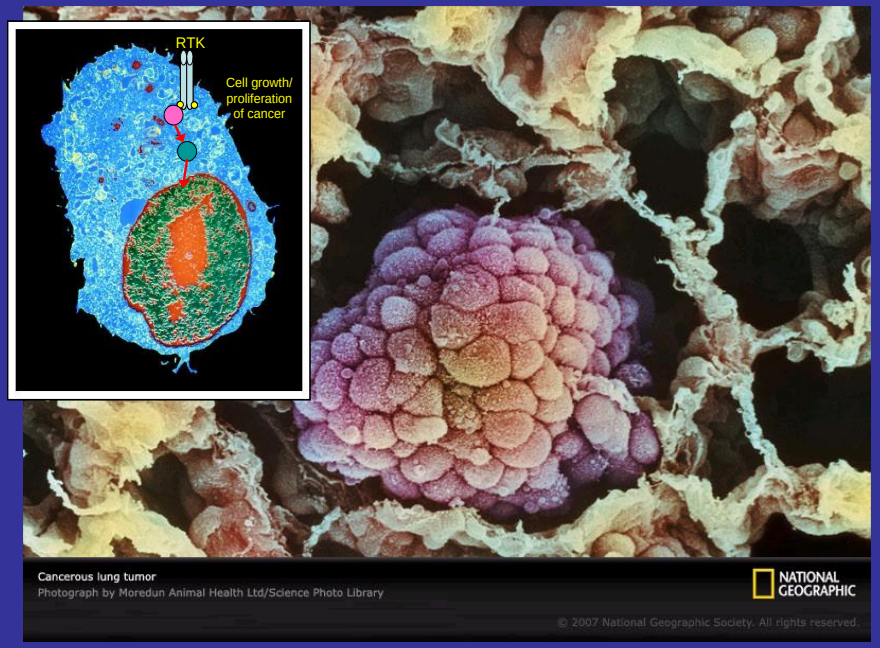
NIH

This way?

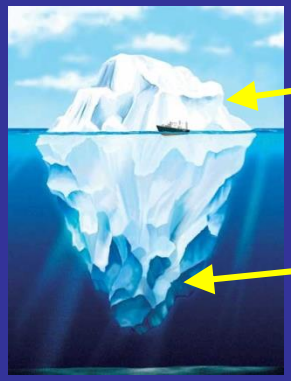
That way?

NSF

Day old proteins from cancer cells



80-85% effort



Abundant/non-specific proteins

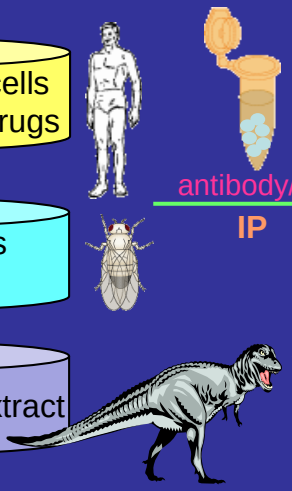
Low level signaling proteins exist at similar levels to dinosaur soft tissue remains

Proteomics Strategy for Identifying Proteins and Modifications

Human cancer cells kinase inhibitor drugs

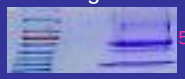
Fly S2R+ cells stimulation

Dinosaur bone extract



antibody/tag

IP



Short gel run

55kDa

Solution

Tryptic digestion

Peptides

enrichment

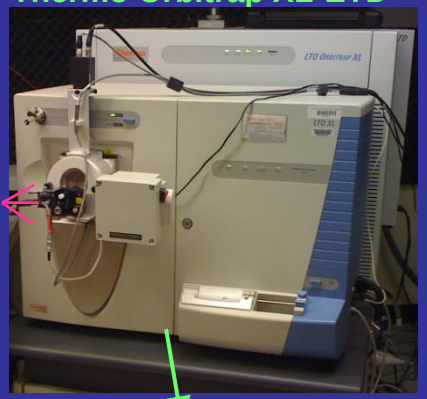
Proxeon EASY-nLC

RP-C_{18}

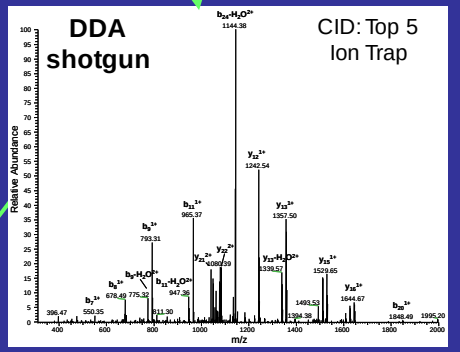
300nL/min

Discovery or Targeted μRP-LC/MS/MS

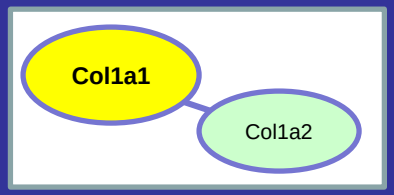
Thermo Orbitrap XL-ETD



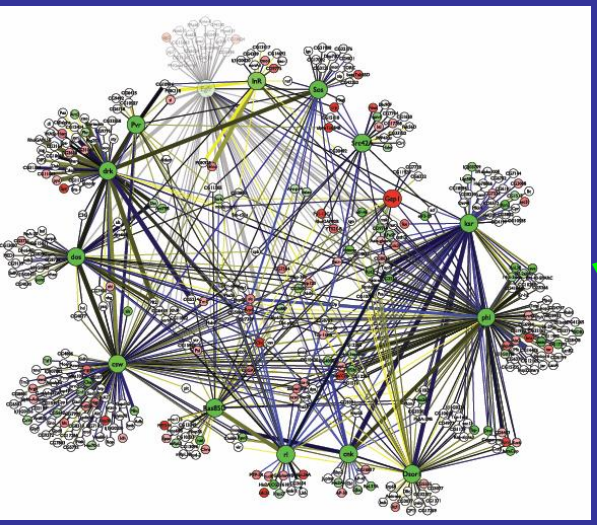
Sequest/ Mascot



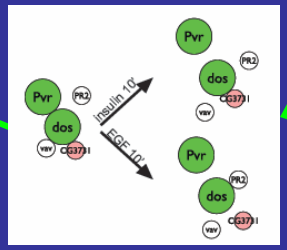
Dinosaur PPI network



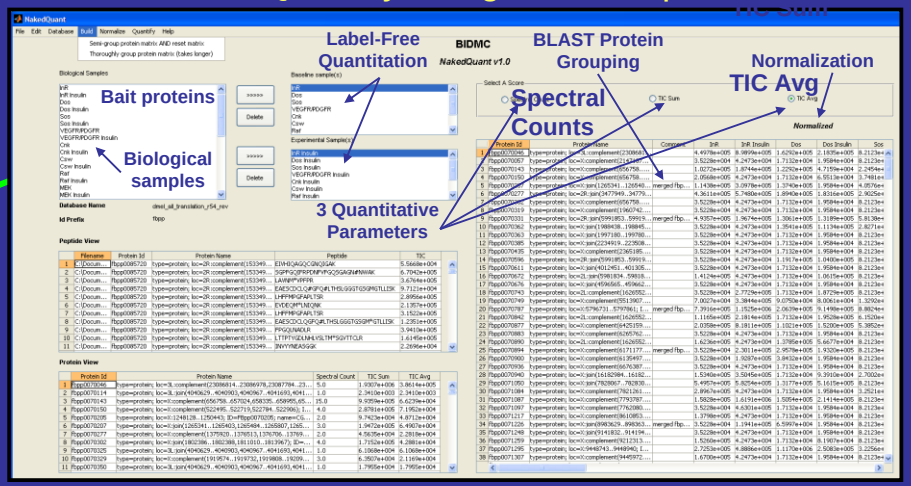
PPI Network



Dynamics (stimulate/drugs)

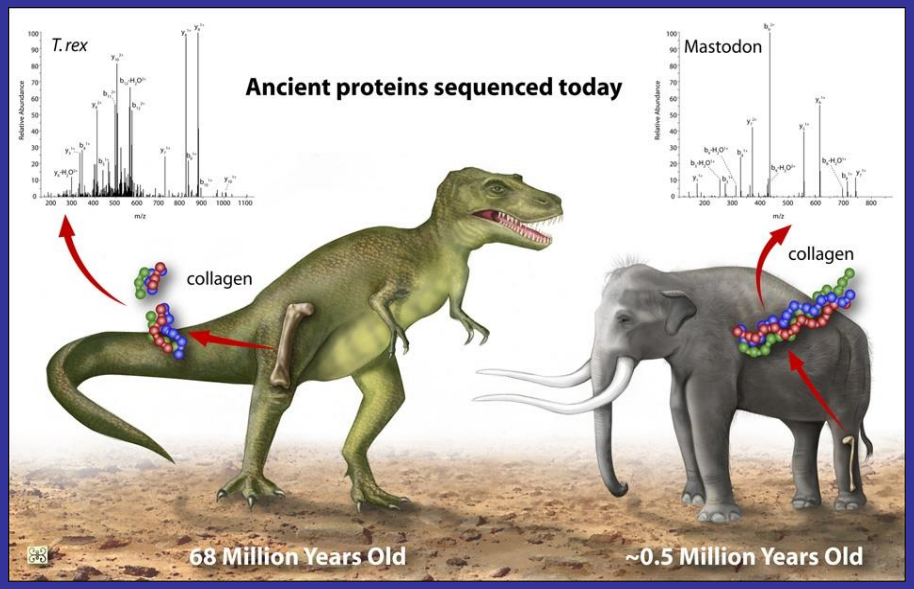


Semi-Quantify using MS/MS spectra



fos·sil·o·mics: *n.* the analysis of the **amino acid sequence**, **post-translational modifications** and **phylogeny** of fossil bone proteins from extinct organisms

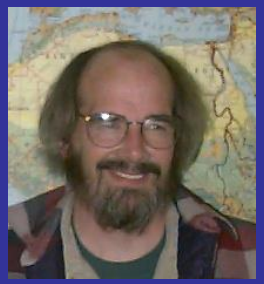
-John Asara
-Noreen Tuross, Harvard University
-Shawnna Donop, formerly of Thermo Fisher Scientific

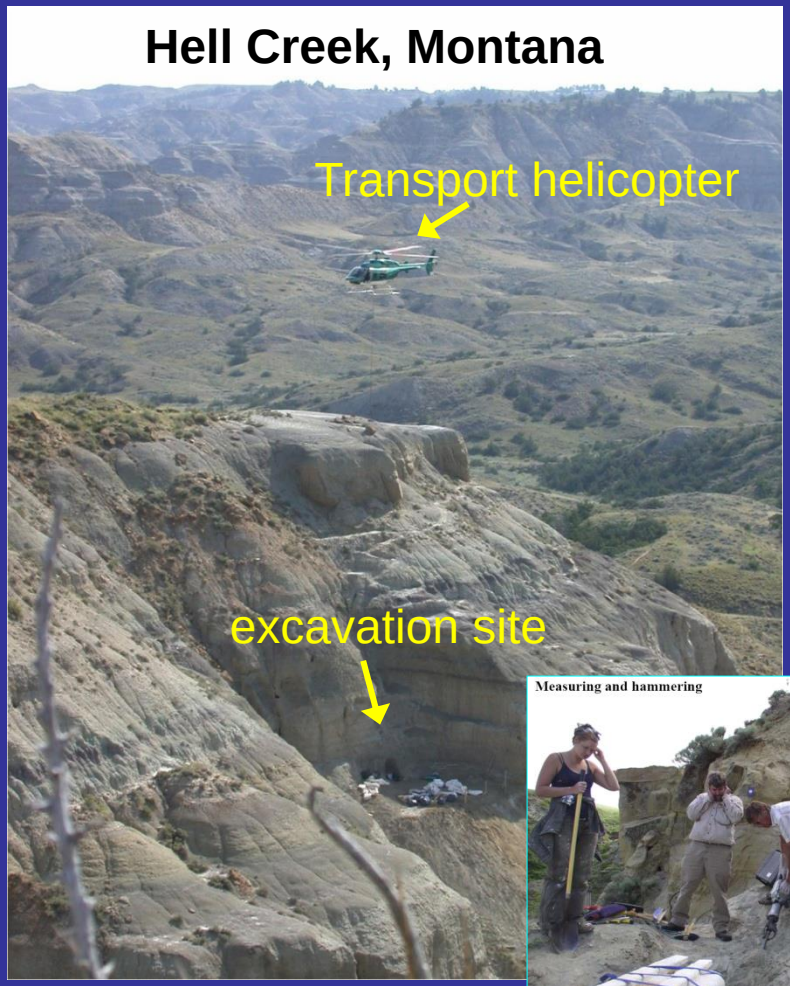


- Most current knowledge of extinct organisms based on the fossil bone record (morphology)
- DNA sequence is not recoverable after ~50-75K years
- Assess the phylogeny of extinct organisms (closest relatives)

65 million years later, *well-preserved* dinosaur fossils are being excavated…

Jack Horner
Museum of the Rockies

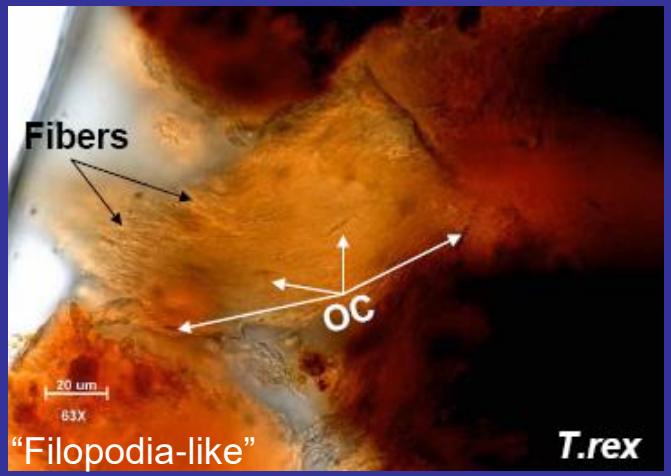




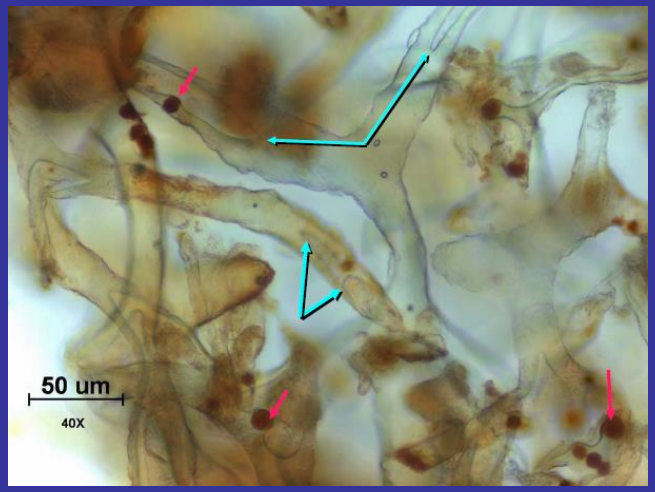
Radiometric dating of surrounding rock dates the *T. rex* fossil to 68 million years

After demineralization, *T. rex* specimen MOR1125 showed amazing soft-tissue preservation

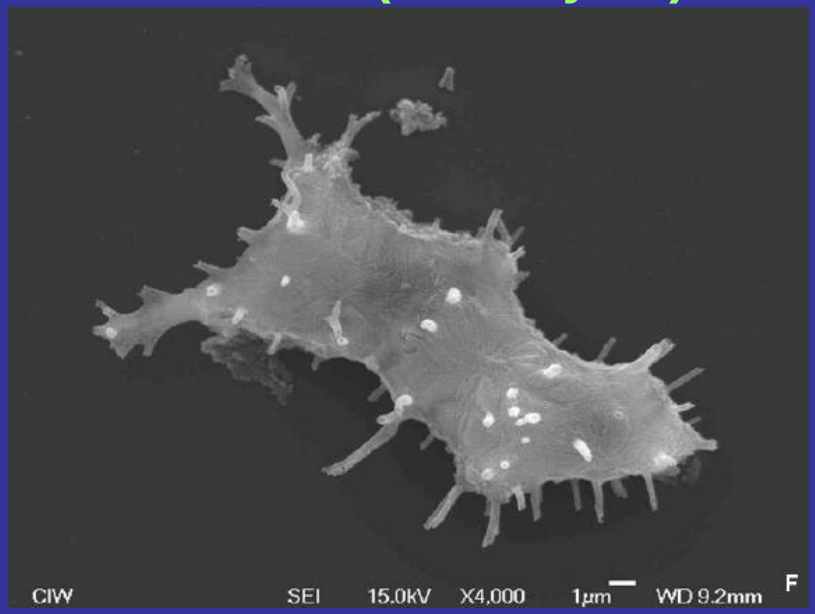
Fibers in cortical bone



Blood vessels in cortical bone



Bone cells (osteocytes)



Mary Schweitzer
NC State University

Schweitzer *et. al*. Science, 2005.

In situ immunohistochemistry of *T. rex* bone

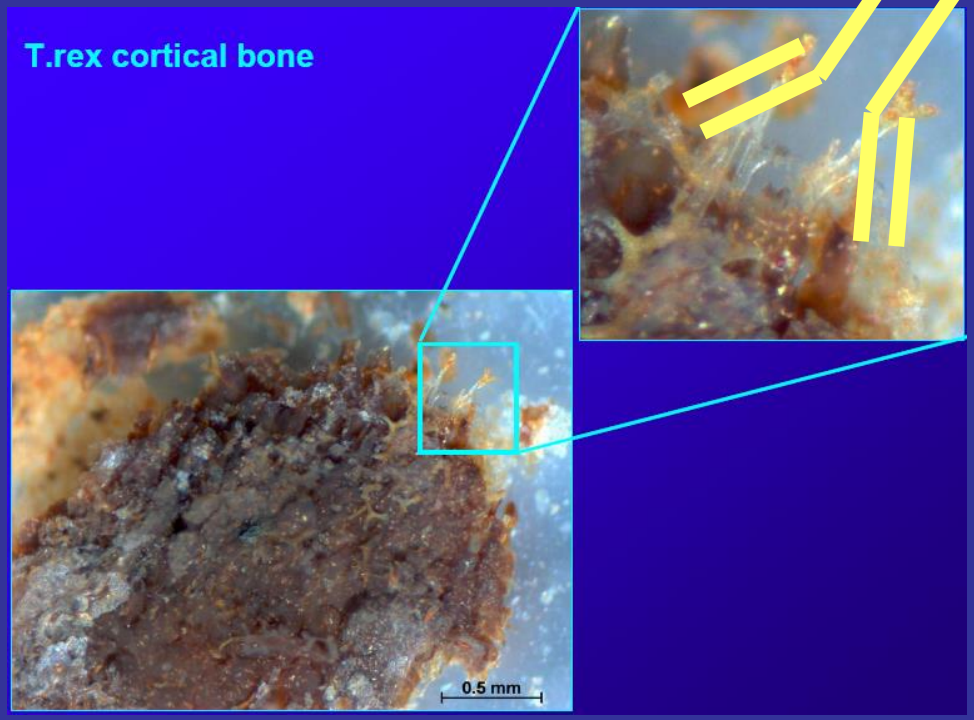
Antibody localized directly to cortical bone

514nm

Secondary antibody

Fluorescent tag

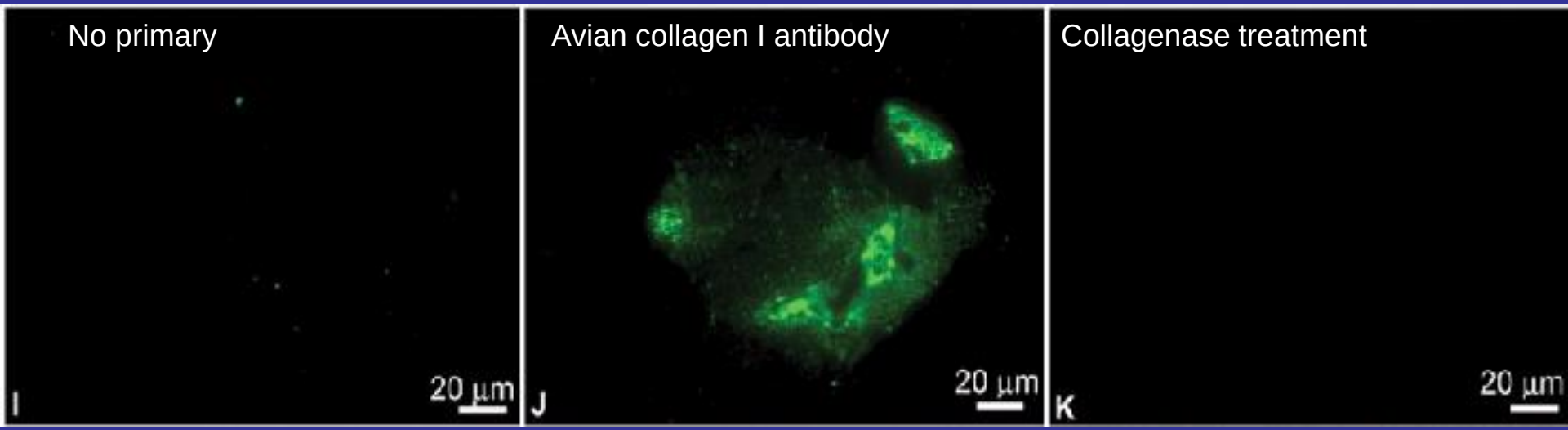
Collagen I antibody



Schweitzer *et. al*. *Science*, 2007.

Demineralized 68 million year old *T. rex* bone reacted to collagen type I antibodies

In situ Immunohistochemistry



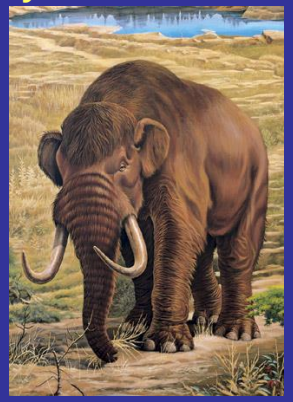
No reactivity against surrounding sediment

Schweitzer *et. al*. *Science*, 2007.

The sensitivity of modern mass spectrometry allows for *many* ancient collagen sequences less than 1 million years

- **Most Ancient Fossil Peptides Match Related Extant Organisms from Protein Database Searches**

~500K year old Mastodon



long bone fragment

Mastodon NCBInr database hits

78 database matches

Protein name	Organism Identity	Number of Peptide Spectra	Amino acid coverage
Collagen alpha-1(I) chain precursor	Dog, bovine, human, chimp	24	20%
Similar to alpha-2 type I collagen	Dog, human	15	10%
similar to collagen alpha 2 type 1	Elephant	12	9%
similar to Collagen alpha 1(IV) chain precursor	Bovine	3	4%
alpha 1 type 1 collagen	Human	2	2%
alpha 1 type II collagen isoform 1	Human	3	4%
Collagen alpha 2(I) chain	Human	4	6%
similar to collagen alpha 1 type 1	Elephant	2	3%
Collagen alpha-1(I) chain	Mouse	2	2%
alpha 1 type II collagen	Newt	2	5%
Similar to alpha-2 type I collagen	Chicken	3	4%
similar to Collagen alpha 1(I) chain precursor	Chimp	2	2%
alpha 1 type 1 collagen	Newt	2	3%

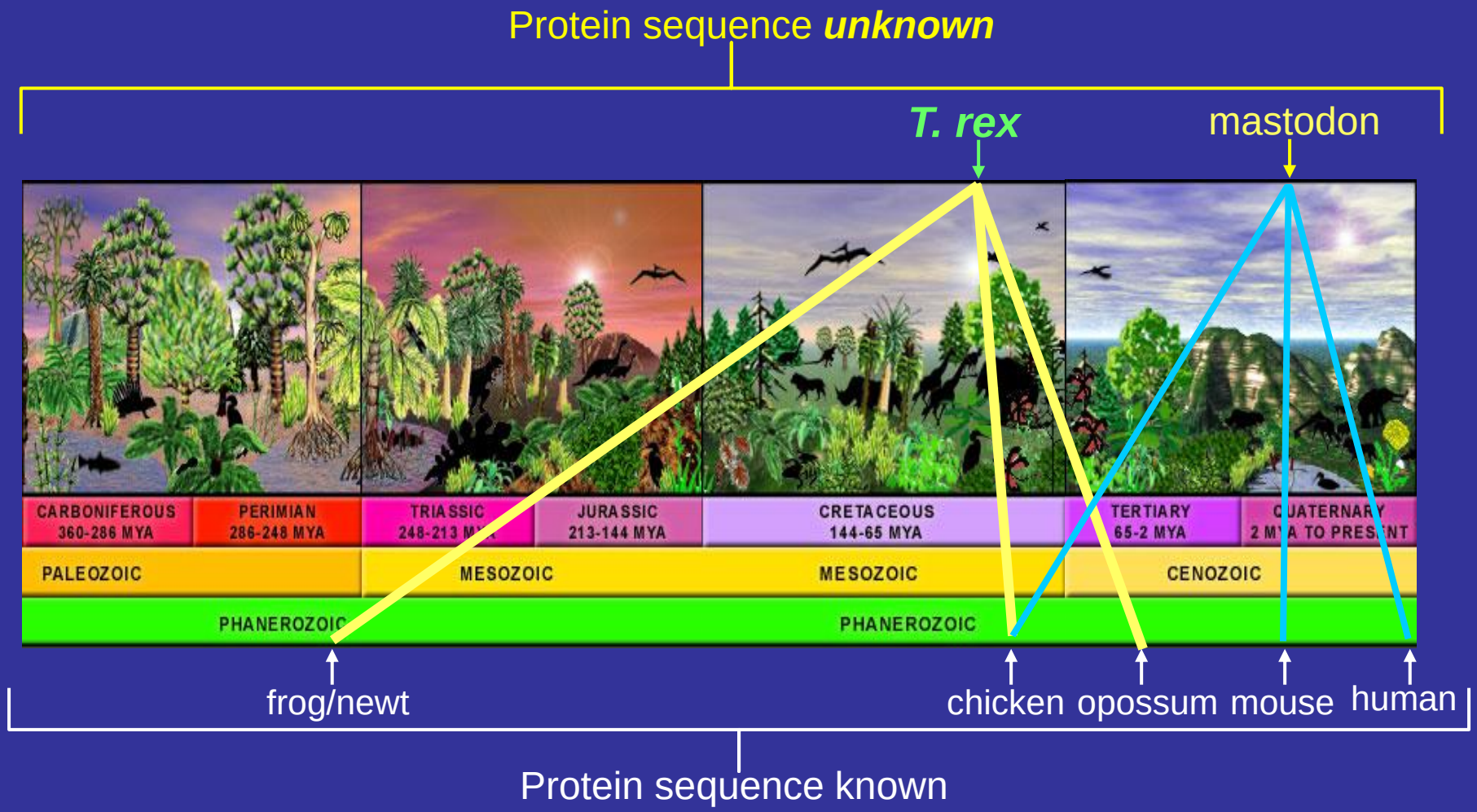
Collagen is highly conserved (human to frog shows 83% identity)
(human to dog shows 97% identity)

LTQ linear Ion Trap

- **For some peptides, *de novo* sequencing is necessary**
  - Found that error tolerant searches were more successful than *de novo* attempts using sequence conservation to our advantage

Bioinformatics Based *De Novo* Sequencing Strategy

•Generate a theoretical (predicted) sequence database to fill the evolutionary gap between organisms of known and unknown sequence



•Produced very minimal results in dinosaur since data was limiting

Example of a *de novo* sequence from ostrich

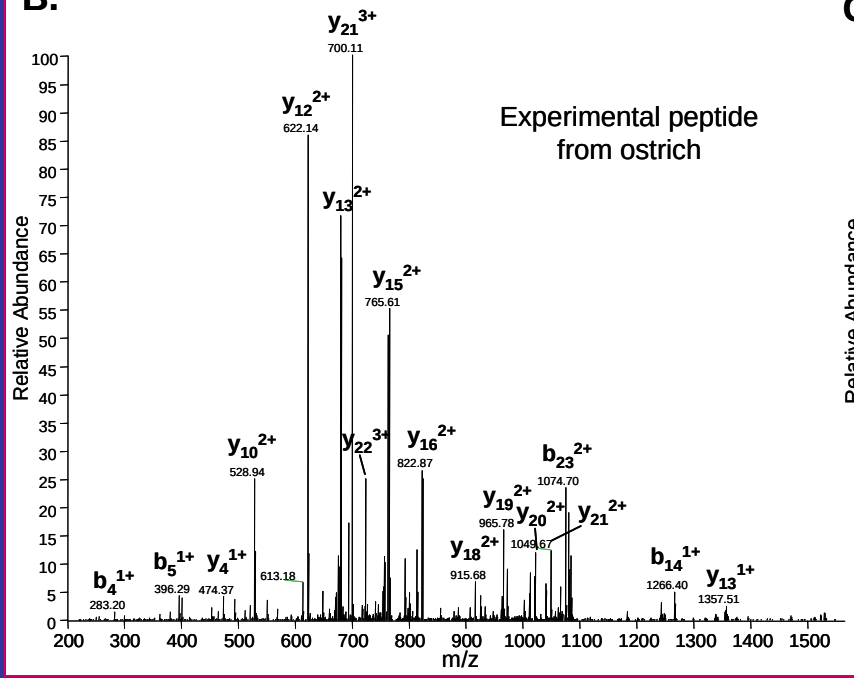
A.

Weighted organism (chicken) produces residue in predicted sequence

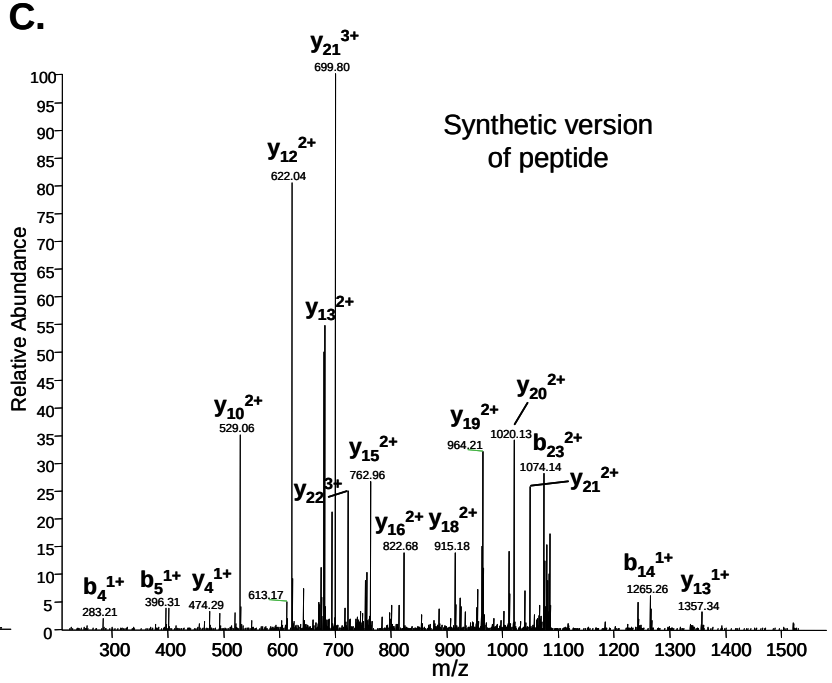
Consensus of at least two organisms produces residue in predicted sequence

P02457 collagen α1t1 Chicken	PGASGPMGPRGPAGPPGKNGDDGEAGKPGRPGQRGPPGPQGARGLPGTAG
CWSC Predicted ostrich peptide	----------GPAGPPGKNGDDGEAGKPGRPGER----------
BAA36973 collagen α1t1 Newt	PGAAGALGPRGLPGPPGKNGDDGESGKPGRPGERGPSGPQGARGLPGTAG
BAA94972 collagen α1t1 Frog	PGASGAMGPRGSSGPPGKNGEDGEAGKPGRPGERGPPGPQGARGLPGTAG

B.



C.



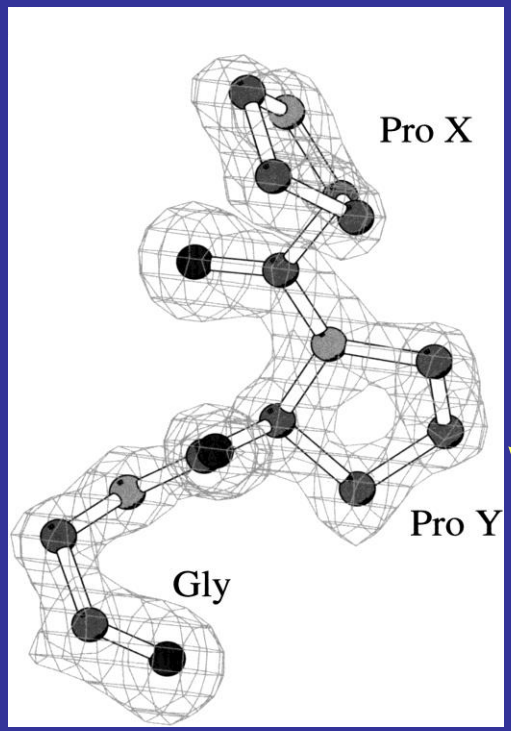
- Produced very minimal results in dinosaur since data was limiting

Asara *et. al*. *Science*, 2007.

Collagen Type I is the most abundant protein (>90%) in bone *and* the most abundant protein in vertebrates

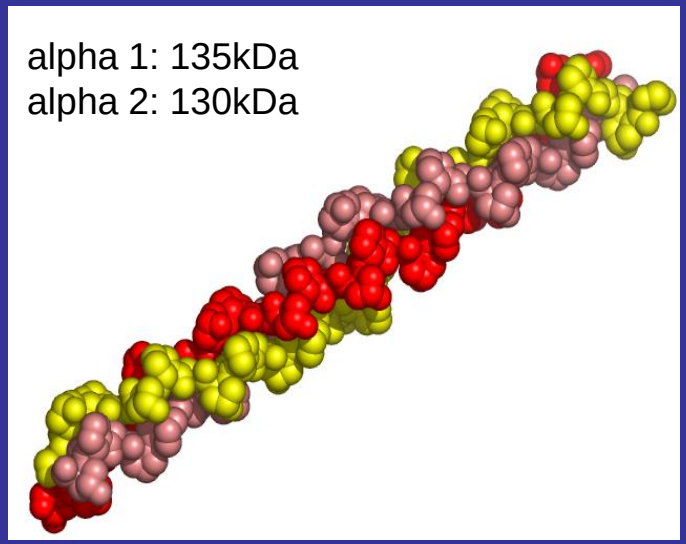
- ~50% of prolines are hydroxylated for structural stability
- ~10% of lysines are hydroxylated for collagen cross-linking

GXY repeat sequence
(33% glycine content)



Hydroxylation of Pro stabilizes the structure

Triple helical structure



- The hydroxylation motif is typically -GX**Y**G- where Y is hydroxylated when Pro

Pekkala *et. al*. *J. Biol. Chem.*, 2004.

Processing of *T. rex* fossil bone extract for LC/MS/MS

***T. rex* bone fragments**

<u>Extract protein</u>

EDTA/acid Demineralization

↓

Grind to powder

↓

Guanidine-HCl extraction

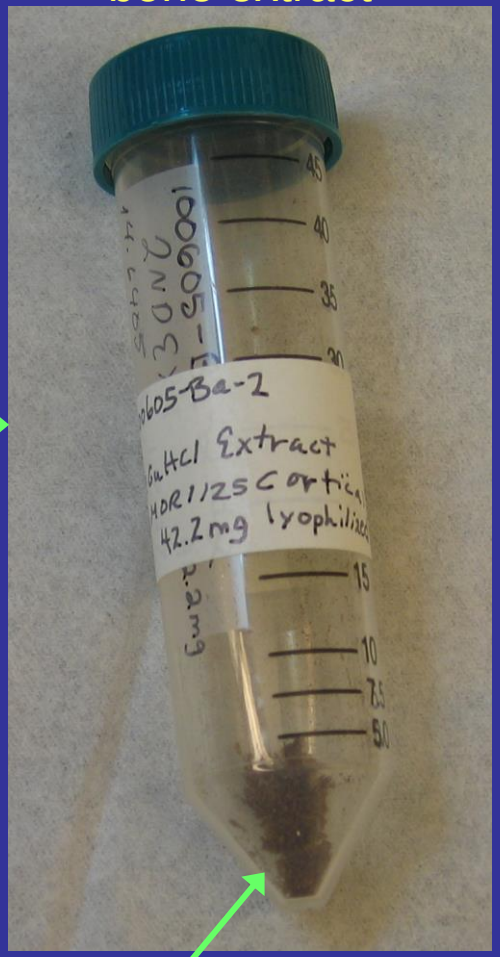
↓

Dialysis

↓

Lyophilization

bone extract



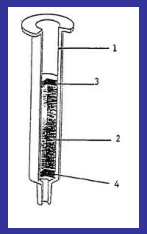
We received this brown stuff

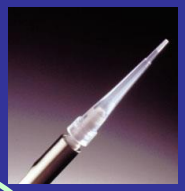
These steps performed in Mary Schweitzer's lab

Needed clean-up prior to Mass Spec!

Successful *T. rex* sample prep strategy

Ion trap for *T. rex* study





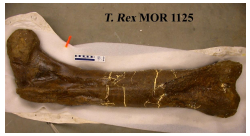
Tryptic digest → C_{18} Sep-Pak → SCX Ziptip → C_{18} Ziptip → μRP-LC/MS/MS

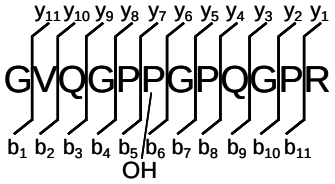
- Tryptic digest: •Bottom-up proteomics
- C_{18} Sep-Pak: •Clean debris •Desalt
- SCX Ziptip: •Enrich peptides •**Cleared brown contaminant!**
- C_{18} Ziptip: •Clean SCX salts •Concentrate peptides
- μRP-LC/MS/MS: •Peptide separation •Fragmentation & sequencing

• 5 bone extractions were performed resulting in 48,216 LC/MS/MS spectra over 1.5 years

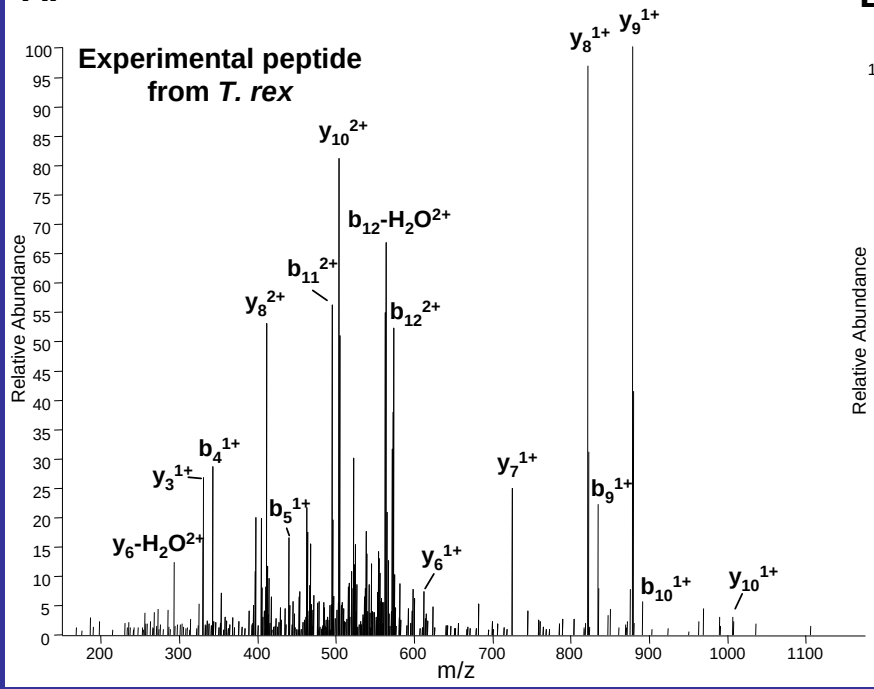
Most spectra did ***not*** yield sequence. . . . .but some did!

MS/MS Spectrum from a 68 Million Year Old Collagen Peptide Acquired with a LTQ Ion Trap

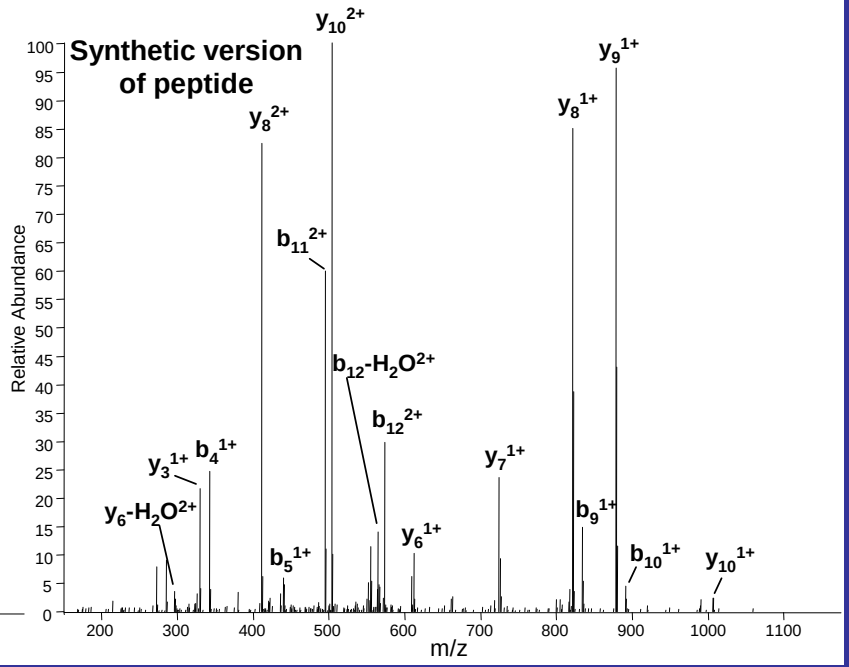




A.



B.

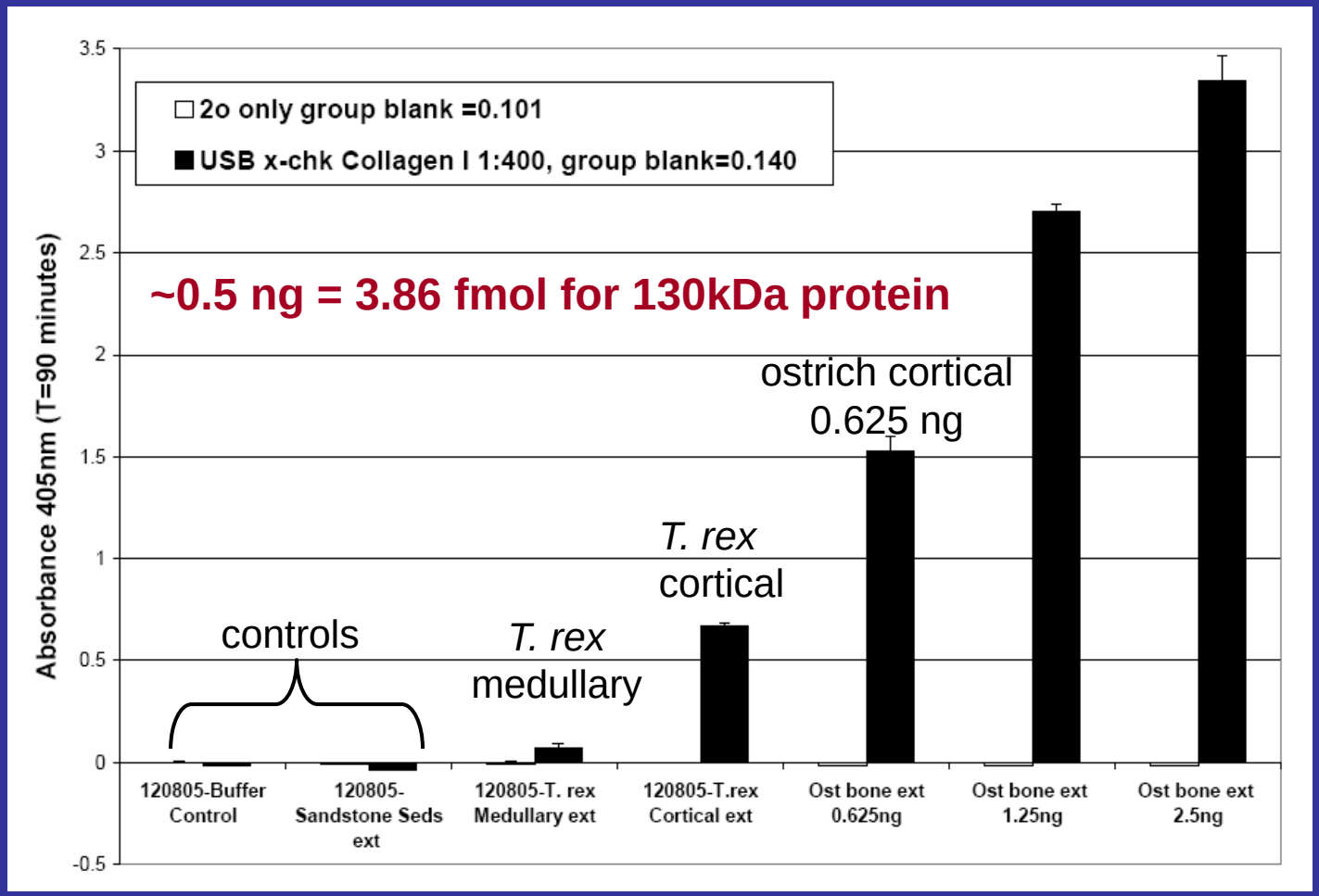


- Only <u>low/sub femtomole</u> levels of peptide available from **~5g** of whole bone producing **~30-40 mg** of bone extract (0.0002%)

- **Protein was never expected to survive past 1 million years!**

Asara *et. al*. *Science*, 2007.

ELISA data suggests collagen exists at sub nanogram levels, consistent with MS peptide sequencing studies



- Sediment and buffer controls produced no collagen signal (no sequences)

Schweitzer *et. al*. *Science*, 2007.

68 Million year old *T. rex* peptide sequences (89 AA)

Peptide sequence	Protein	Organism identity by BLAST
GATGAP*GIAGAP*GFP*GAR	Collagen α1t1	Chicken and frog
GSAGPP*GATGFP*GAAGR	Collagen α1t1	Multiple organisms
GVQGPP*GPQGPR	Collagen α1t1	Chicken and opossum
GLPGESGAVGPAGPIGSR	Collagen α2t1	Chicken
GVVGLP*GQR	Collagen α1t1	Multiple organisms
GAPGPQGPSGAP*GPK	Collagen α1t1	Novel

Chicken was top BLAST match

* Hydroxylation PTM
- No representation by a single organism argues against contamination
- Sediment surrounding bone showed no collagen

Difficulties in sequencing collagen with ion trap

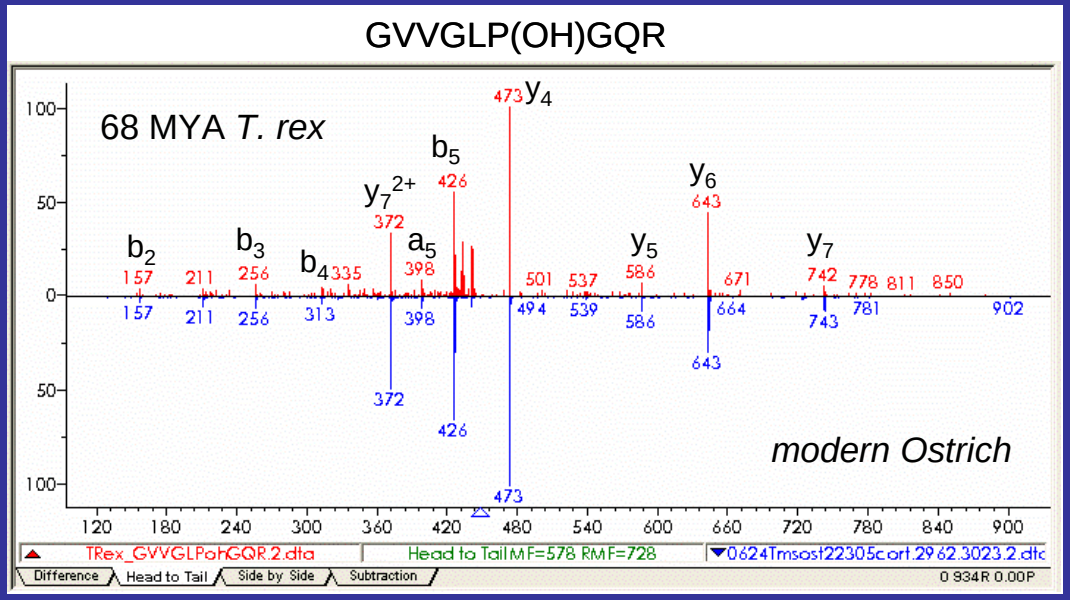
-P(OH)- ↔ -I/L-	**113 Da (0.0364 Da difference)**
-AP(OH)- ↔ -SP-	87 Da (no mass difference)
-G(OH)A- ↔ -GS-	87 Da (no mass difference)
Adjacent P for hydroxylation	+16 Da (no mass difference)

Asara *et. al.,* Letter to *Science*, 2007.

Rigorous statistical validation of *T. rex* collagen sequence using Mascot for phylogenetic analysis

Spectrum #	m/z(obsd)	Mr(Calc)	Mass error	Rank	Mascot score	Expectation value	Validation	Peptide
13032	450.04	897.50	0.56	1	22.1	49	ostrich peptide	GVVGLP(OH)GQR
19231	581.59	1161.59	-0.42	1	32.0	5.5	synthetic peptide	GVQGPP(OH)GPQGPR
28920	787.10	1571.77	0.42	1	42.4	0.44	multiple matches	GATGAP(OH)GIAGAP(OH)GFP(OH)GAR
29032	790.06	1577.82	0.28	1	46.8	0.16	Search stats.	GLPGESGAVGPAGPIGSR
26727	730.53	1458.69	0.36	1	62.3	0.0048	Search stats.	GSAGPP(OH)GATGFP(OH)GAAGR
22865	645.76	1289.64	-0.12	1	71.1	0.00067	Search stats.	GAPGPQGPSGAP(OH)GPK

48,216 total spectra were collected, 2 peptides fall below 5% False Positive Rate



Spectral comparison versus >200,000 peptide spectra using MS Search 2.0 (NIST); scored 934 of a possible 999; next best score was 89.

Lam *et. al.*, *Proteomics*, 2007.

John Cottrell, Matrix Science

Asara *et. al. Science*, 2008, *Response*.

Phylogenetic analysis by comparison with 21 extant organisms

Fish


Danio rerio


Oncorhynchus mykiss


Raja kenojei


Paralichthys olivaceus

Amphibians


Cynops pyrrhogaster

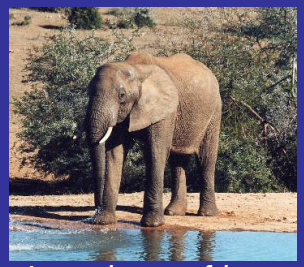

Xenopus tropicalis


Xenopus laevis

Mammals


Macaca mulatta


Echinops telfairi


Loxodonta africana


Homo sapiens (my sons)


Canis lupus familiaris


Pan troglodytes


Mus musculus


Rattus norvegicus


Bos taurus


Monodelphis domestica

Reptiles

LC/MS/MS


Anolis carolinensis


Alligator mississippiensis

Birds

LC/MS/MS


Gallus gallus

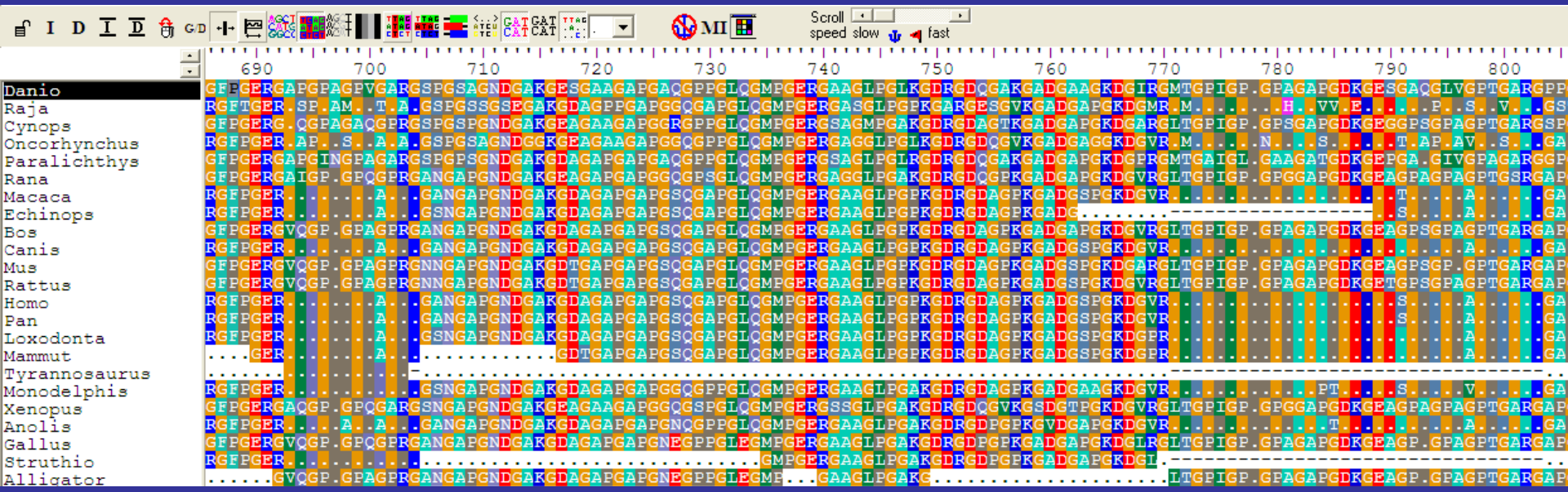

Struthio camelus

Critical organisms

Alignment of Collagen Type I α1 and α2 Sequences:

21 extant organisms

2 extinct organisms



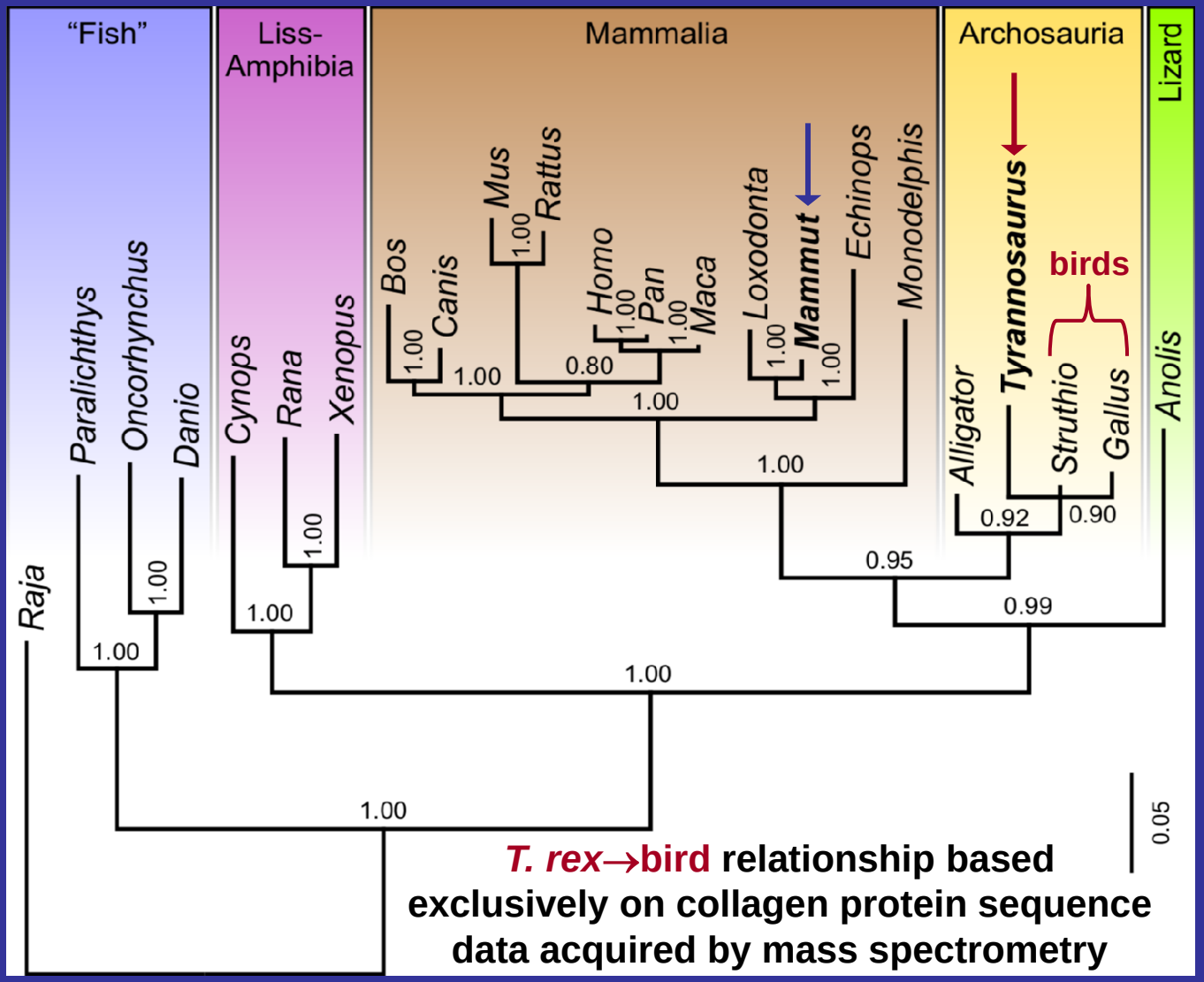
A small portion of the ClustalX alignment is shown

Phylogenetic Analysis: **Bayesian - samples tree millions of times**

BayesPhylogenies

Chris Organ, Harvard University

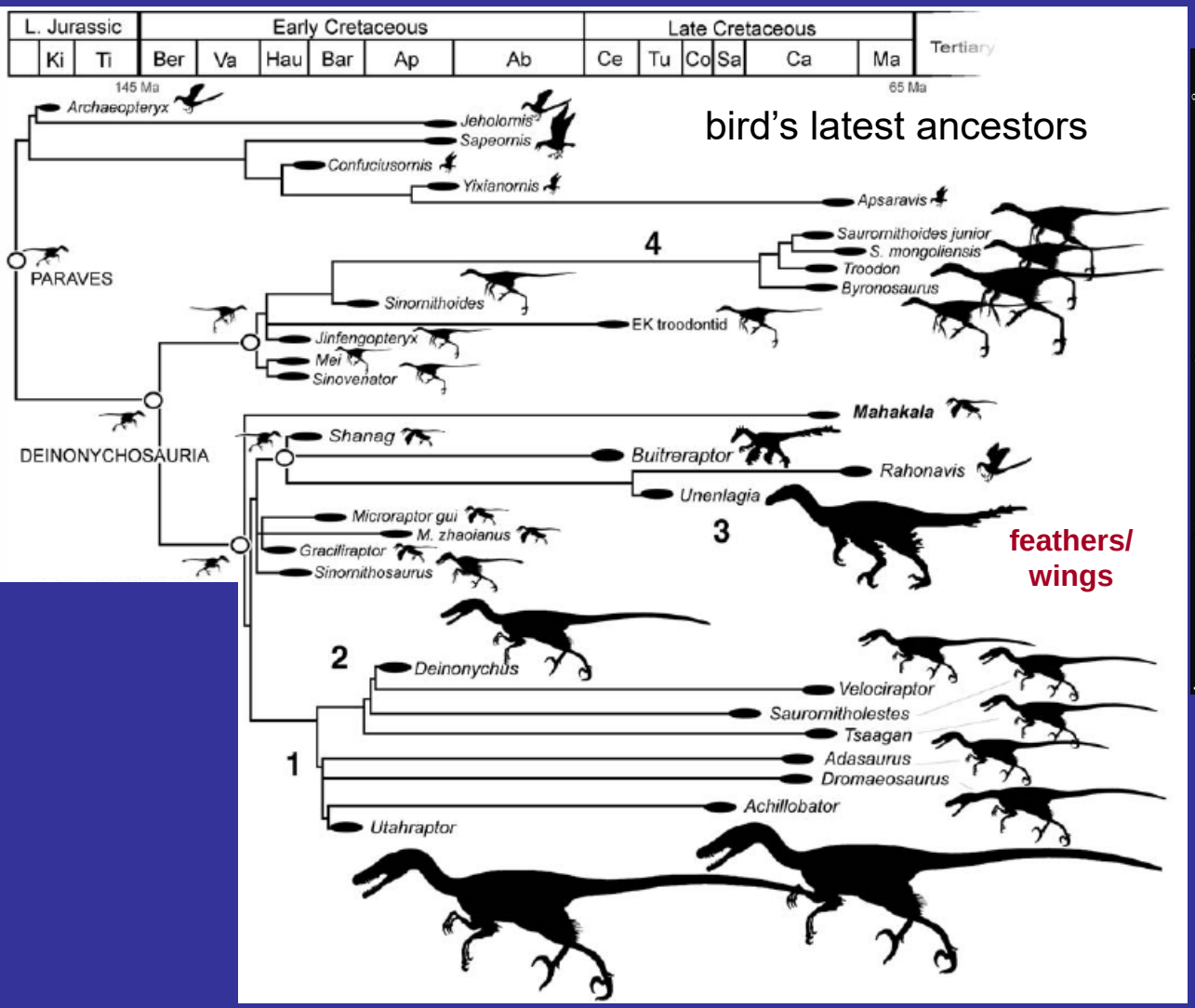
Phylogenetic Tree of *T. rex* from Sequence Data

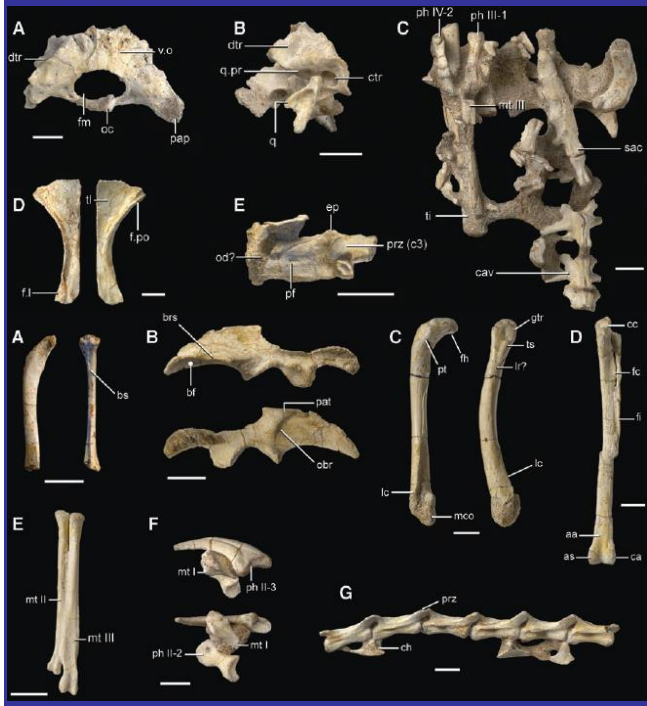


- Bayesian tree shows very good organization of both modern and extinct organisms
- Validated using several algorithms (maximum likelihood, parsimony, and distance)

Organ *et. al., Science*, 2008.

Overwhelming evidence of *Dinosaur* → *Bird* evolution based on the fossil record





- Previous data based on bones (morphology), *not* sequence data

Turner *et. al., Science* 2007.

Controversy begins (as expected) **……**

- Bacterial biofilm ?
  - Kaye, et. al., 2008, PLoS ONE
- Statistical significance of collagen peptide hits ?
  - Pevzner et. al., 2008, Technical comment in *Science*
- Contamination ?
  - Still waiting – a couple of unsuccessful attempts (Fitzgibbon & McIntosh)

Public release of 48,216 *T. rex* raw MS/MS spectra to PRIDE database

-We previously only released spectra supporting the published sequences

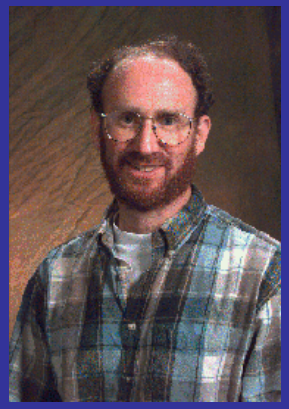
Reanalysis of publicly released 48, 216 *T. rex* raw spectra by independent groups…..

A tale of two computer scientists..

Bad cop

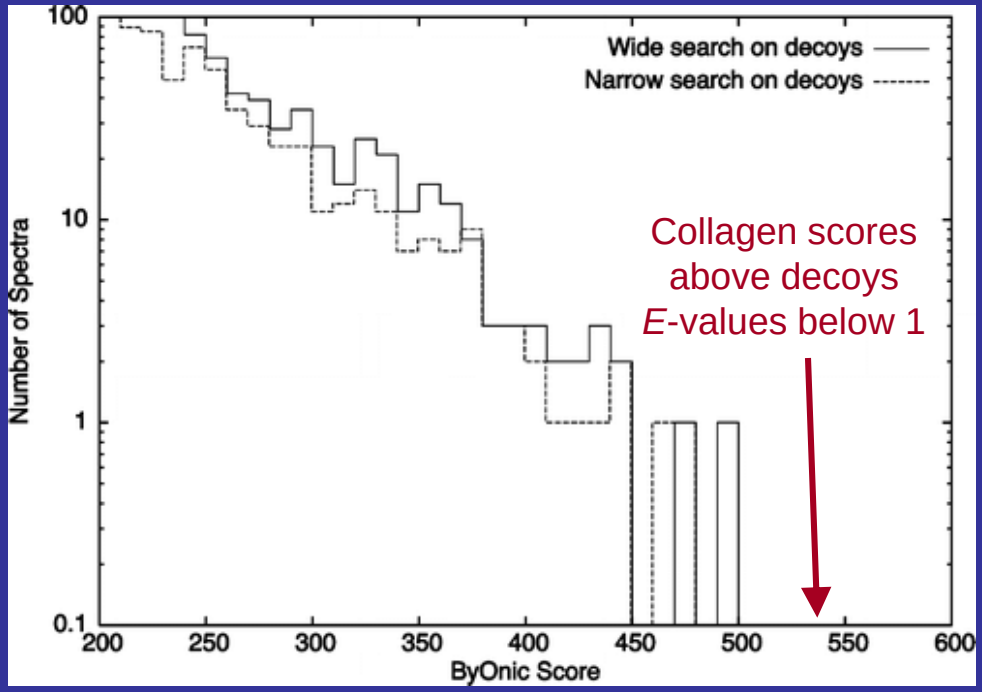
Pavel Pevzner
UCSD

Good cop



Marshall Bern
Palo Alto Research Center

Bern showed statistically significant collagen using ByOnic



"Assuming statistical independence of distinct peptides, the identification of bird-like collagen at the protein level is clearly significant."

Bern, M. et. al. *J Proteome Res.* 2009 Sep;8(9):4328-32.

Reanalysis of *T. rex* spectra by Bern *et. al.* confirmed our findings……

Table 2. The Spectra Matching Collagen and Hemoglobin[a]

No single organism represented

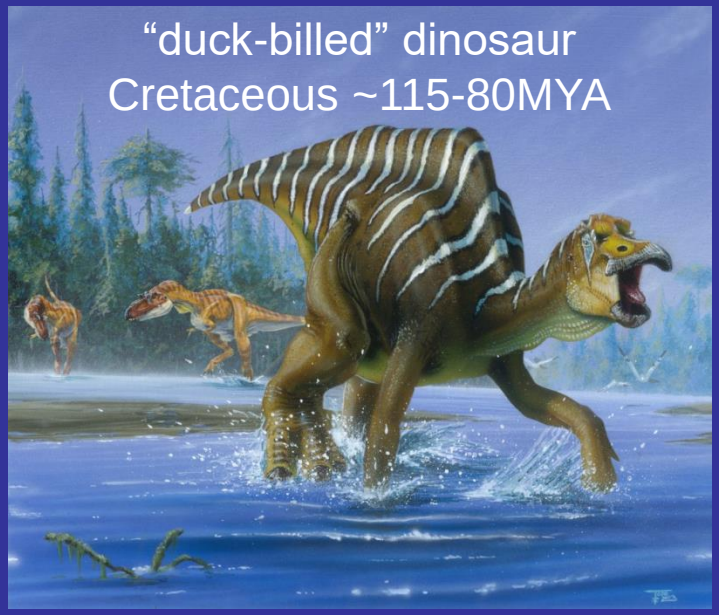
scan	peptide	organisms	score
0130Tmsdinoedtac.2295.2295.2.dta	R.GLAGPQGPR.G	Collagen alpha-2(I) (fragments), *G. Gallus*	258
0130Tmsdinoedtac.2357.2357.2.dta	R.GAPGPQGP[+16]AGAPGP[+16]K.G	Collagen alpha-1(I), *C. pyrrhogaster, T. rex*	334
0607Tmscorh2x.997.997.2.dta	R.GAPGPQGPSGAP[+16]GPK.X	Collagen alpha-1(I), *T. rex* **Novel sequence**	526
0130Tmsdinoedtac.2736.2736.2.dta	R.GPP[+16]GESGAAGPTGPIGSR.G	Collagen alpha-2(I), *H. sapiens, B. taurus*	257
0130Tmsdinoedtac.2784.2784.2.dta	R.GSAGPP[+16]GATGFP[+16]GAAGR.V	Collagen alpha-1(I), *G. Gallus*	509
0421Tjadinocortzipr.3000.3000.2.dta	R.GSAGPP[+16]GATGFP[+16]GAAGR.V	Collagen alpha-1(I), *G. Gallus*	404
0421Tjadinomedzip.3021.3021.2.dta	R.GSAGPP[+16]GATGFP[+16]GAAGR.V	Collagen alpha-1(I), *G. Gallus*	435
0628Tmsmor1125v.3341.3341.2.dta	R.GSAGPP[+16]GATGFP[+16]GAAGR.V	Collagen alpha-1(I), *G. Gallus*	387
0130Tmsdinoedtac.3160.3160.2.dta	K.GATGAP[+16]GIAGAP[+16]GFP[+16]GAR.G	Collagen alpha-1(I), *G. Gallus*	309
0419Tjatrexscxc18zip.3552.3552.2.dta	K.GATGAP[+16]GIAGAP[+16]GFP[+16]GAR.G	Collagen alpha-1(I), *G. Gallus*	435
0421Tjadinomedzip.3451.3451.2.dta	K.GATGAP[+16]GIAGAP[+16]GFP[+16]GAR.G	Collagen alpha-1(I), *G. Gallus*	536
0628Tmsmor1125v.3951.3951.2.dta	K.GATGAP[+16]GIAGAP[+16]GFPGA[+16]R.G	Collagen alpha-1(I), *G. Gallus**	327
0130Tmsdinoedtac.2928.2928.2.dta	R.GL[-16]P[+16]GESGAVGPAGPIGSR.G	Collagen alpha-2(I) (fragments), *G. Gallus**	265
0421Tjadinocortzipr.2931.2931.2.dta	R.GVVGLP[+16]GQR.G *	Collagen alpha-1(I), *G. gallus, H. sapiens*	361
0421Tjadinocortzipr.3057.3057.2.dta	R.GEP[+16]GPAGLP[+16]GPAGER.G	Collagen alpha-1(I), *G. Gallus*	324
0607Tmscorh2x.3432.3432.2.dta	K.GATGAP[+16]GIAGAP[+16]GFP[+16]GAR.G	Collagen alpha-1(I), *G. Gallus*	389
0607Tmscorh2x.3435.3435.2.dta	K.GATGAP[+16]GIAGAP[+16]GFP[+16]GAR.G	Collagen alpha-1(I), *G. Gallus*	447
0628Tmsmor1125v.2855.2855.2.dta	R.GVQGPP[+16]GPQGPR.G *	Collagen alpha-1(I), *G. Gallus*	410
0628Tmsmor1125v.2963.2963.2.dta	R.GP[+16]PGSS[-16]GSTGK.D	Collagen alpha-1(I), *C. pyrrhogaster**	250
0628Tmsmor1125v.4350.4350.2.dta	K.GAAGLPGV[+16]AGAP[+16]GLPGP[+16]R.G	Collagen alpha-2(I) (fragments), *G. Gallus**	292
112905Tmsd5.1489.1489.2.dta	R.P[+16]GC[+57]P[+16]GPMGEK.G	Collagen alpha-4(IV) chain, *M. musculus*	345
	* **Low-scoring by Mascot**		
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

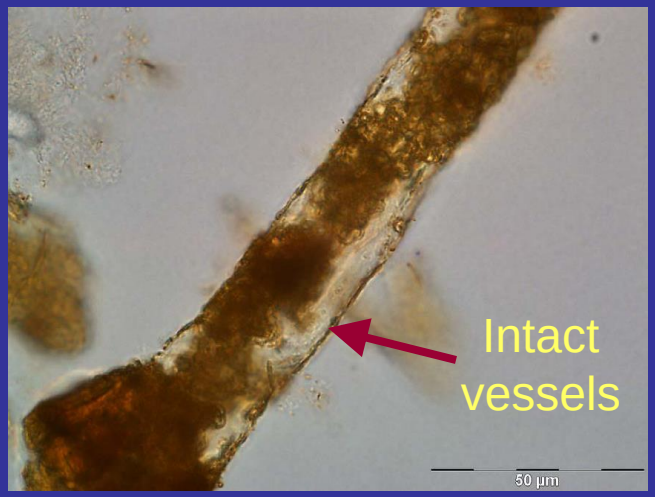
"In summary, we find nothing obviously wrong with the *T. rex* mass spectra: the identified peptides seem consistent with a sample containing old, quite possibly very ancient, bird-like bone, contaminated with only fairly explicable proteins."

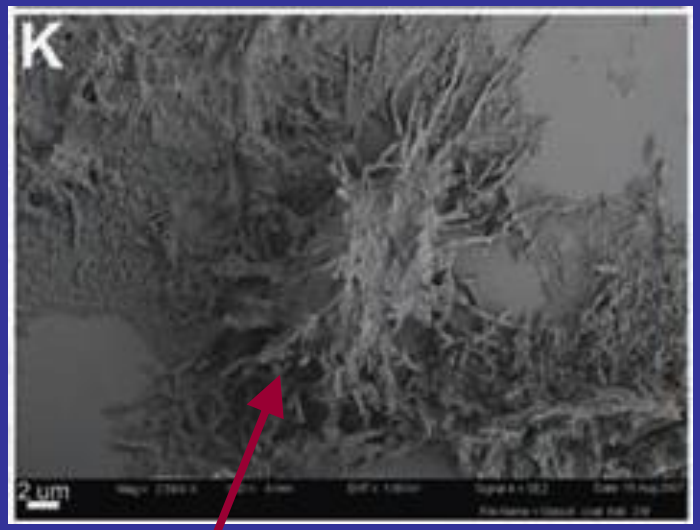
- 7,085 spectra from sediment surrounding *T. rex* bone showed *no* collagen or hemoglobin – **argues against contamination**

Bern, M. et. al. *J Proteome Res.* 2009 Sep;8(9):4328-32.

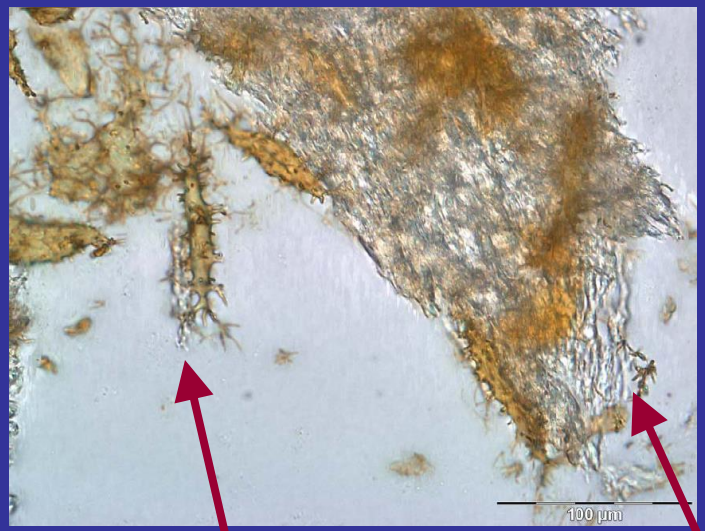
DINO #2: 80 Million year old *Brachylophosaurus canadensis (Hadrosaur)* fossil from Judith River Formation in Montana also preserved soft tissue







SEM of osteocyte in long filopodia



Bone cells in collagen

Collagen fibers

Schweitzer *et. al.*, *Science*, 2009.

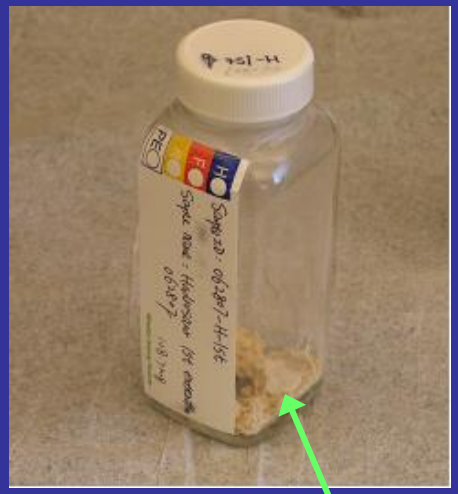
Hadrosaur LC/MS/MS Strategy

Difficulties in sequencing collagen

-P(OH)- ↔ -I/L-	**113 Da (0.0364 Da difference)**
-AP(OH)- ↔ -SP-	87 Da (no mass difference)
-G(OH)A- ↔ -GS-	87 Da (no mass difference)
Adjacent P for hydroxylation	+16 Da (no mass difference)

Acquired high mass accuracy MS!

80MYA Hadrosaur bone extract



Clean and fluffy

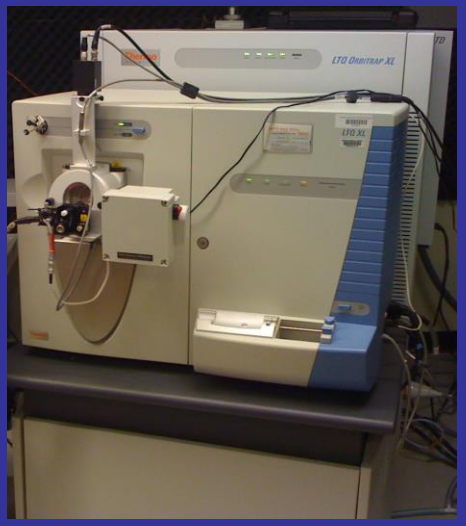
Tryptic Digest C_{18}/SCX Ziptip

Proxeon EASY-nLC

μLC/MS/MS

C_{18} 75μm x 10cm

Thermo LTQ Orbitrap XL

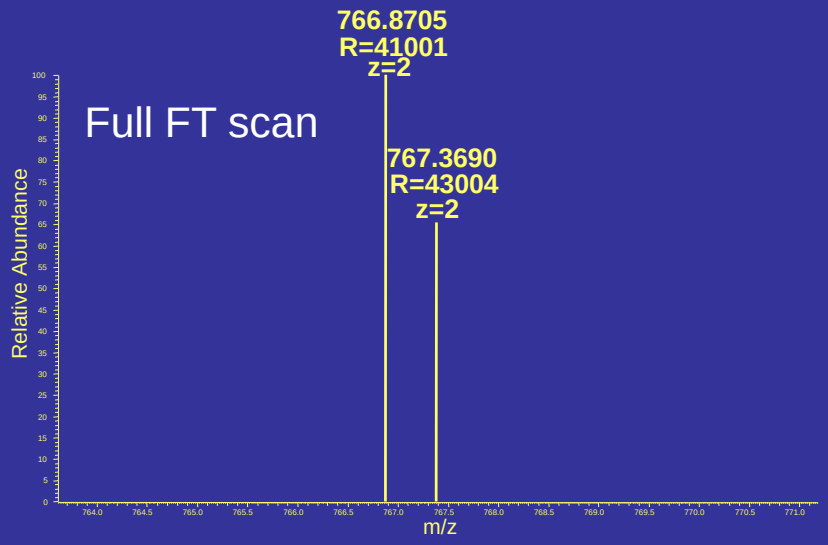


The Orbitrap XL FT full scan resolves difficult sequence calls

Collagen alpha 1 type I Hadrosaur peptide

Original Sequest call: GETGPAGPAGP**I**GPAGAR
MH+ 1532.7450 Δmass: **0.0366** XCorr: **2.9997** Sp: **692.3** Sf: **0.65** P: **42**

24 ppm mass difference?
Can't be correct

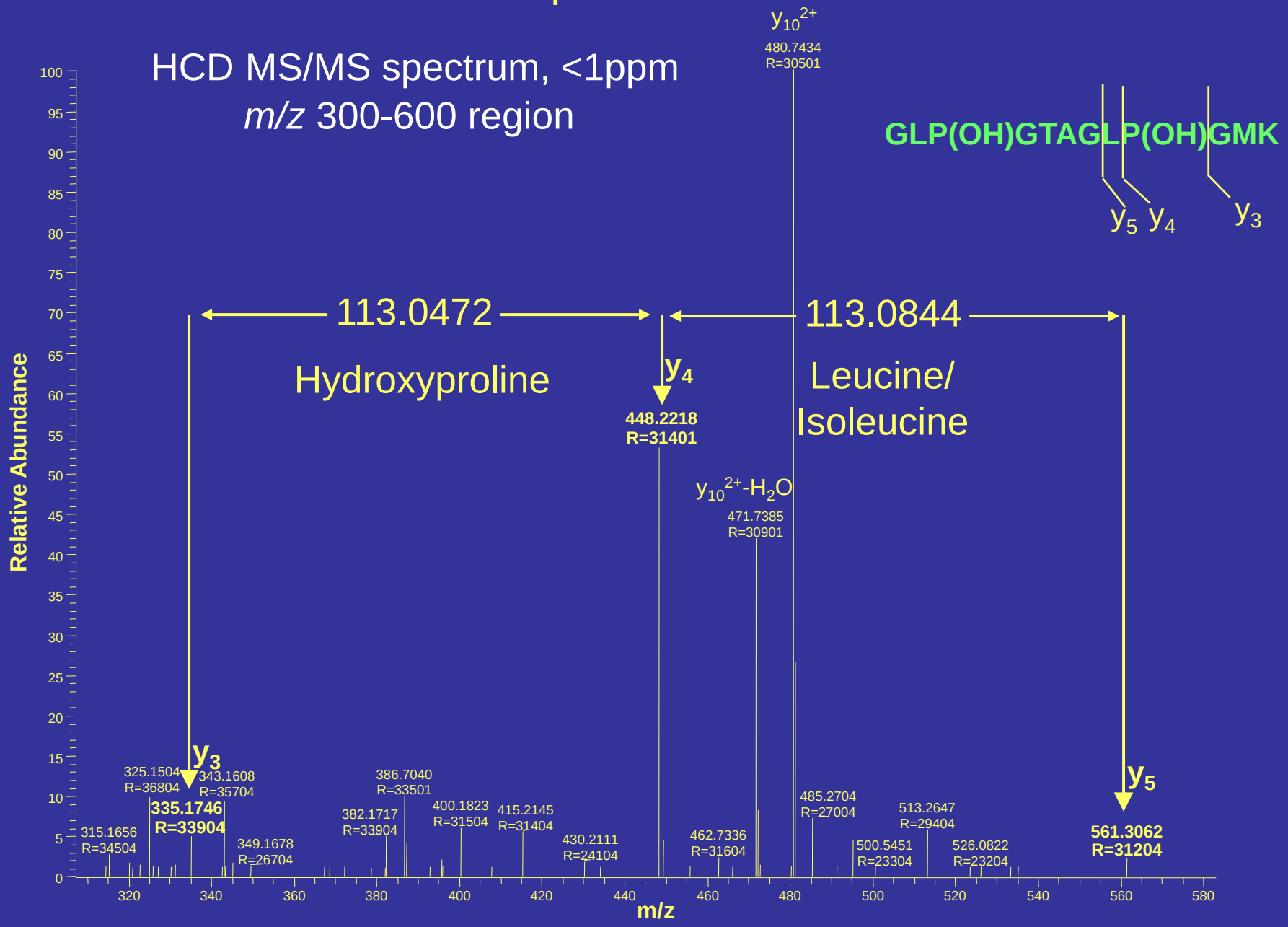


- Typically, <2 ppm mass accuracy

Correct Sequest call: GETGPAGPAGP**P(OH)**GPAGAR
MH+ 1532.7086 Δmass: **0.0002** XCorr: **2.9997** Sp: **692.3** Sf: **0.65** P: **42**

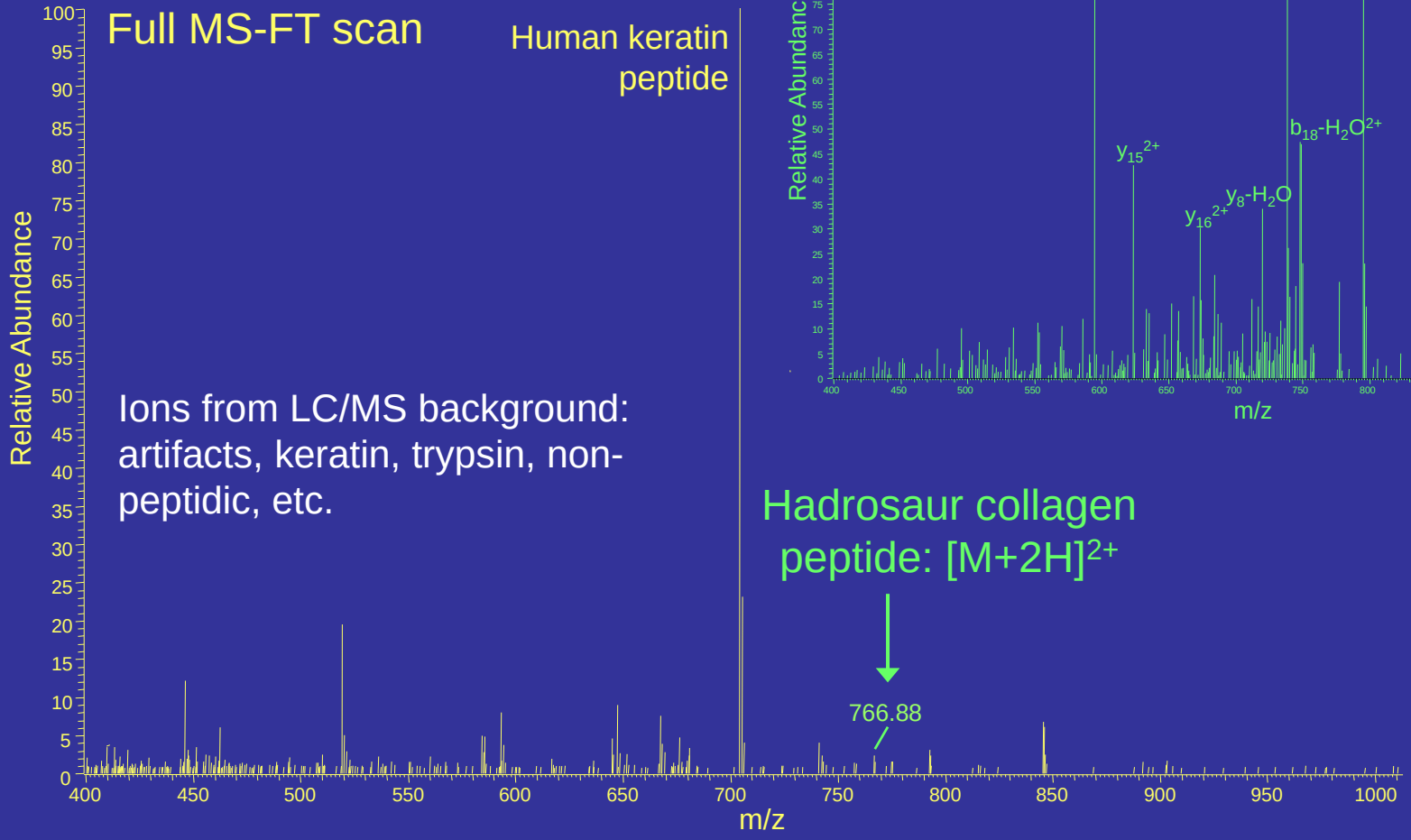
Represents 0.1 ppm mass accuracy; identical Sequest® scores

HCD collision cell on LTQ-Orbitrap XL can resolve *close* sequence calls



From ~35-40 mg of extract, how much collagen signal is obtained from multi-million year old bones?

- *Not much* (low/sub femtomole)

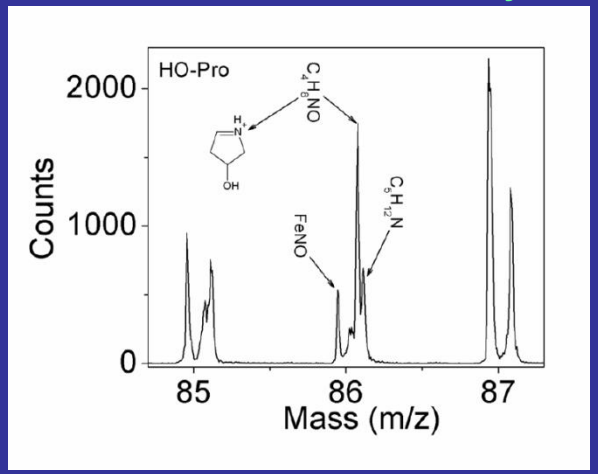


In situ immunohistochemistry from 80 million year old Hadrosaur (*Brachy*) bone tissue

Collagen I antibody | Secondary Ab, no primary | Digested with collagenase

No reactivity against surrounding sediment

TOF-SIMS: surface analysis



Recep Avci

Hydroxyproline is collagen-specific

Can't be bacterial biofilm

-bacteria do not produce collagen

Schweitzer…..Asara, *Science* 2009.

Brachylophosaurus sequences acquired by Ion trap and Orbitrap mass spectrometry

8 collagen sequences, 149 Amino Acids

m/z (obsd)	Mr (calc)	Mass error	Instrument	rank	Mascot score	Mascot expectation value	Sequest Xcorr	Validation		Peptide sequence	Protein	BLAST sequence identity
960.487	2878.421	0.0047*	Orbitrap	1	40.0	0.59	4.53	Search stats; ostrich peptide		GLTGPIGPP(OH)GPAGAP(OH) GDKGEAGPSGPPGPTGAR	Collagen α1(1)	Ostrich and mammals
730.740	1458.685	0.7793	Ion trap	1	73.7	0.00027	3.99	Search stats	*	GSAGPP(OH)GATGFP(OH) GAAGR	Collagen α1(1)	*T. rex*, chicken, and mammals
786.901	1571.769	0.0180*	Orbitrap	1	37.2	0.84	3.13	Search stats; synthetic peptide		GATGAP(OH)GIAGAP(OH) GFP(OH)GAR	Collagen α1(1)	*T. rex*, chicken, alligator, and amphibia
766.877	1531.738	0.0005	Orbitrap	1	52.7	0.023	2.70	Search stats		GETGPAGPAGPP(OH)GPAGAR	Collagen α1(1)	Chicken
582.160	1161.589	0.7164	Ion trap	1	65.8	0.0015	2.48	Search stats; synthetic peptide	*	GVQGPP(OH)GPQGPR	Collagen α1(1)	*T. rex*, chicken, alligator, and opossum
653.824	1305.631	0.0013	Orbitrap	1	56.9	0.012	2.63	Search stats	*	GPSGPQGPSGAP(OH)GPK	Collagen α1(1)	Chicken, alligator, rat, and opossum
805.875	1609.734	0.0012	Orbitrap	1	40.5	0.54	2.32	Search stats; synthetic peptide		GSN(deam)GEP(OH)GSAGPP (OH)GPAGLR	Collagen α2(1)	Chicken and alligator
789.898	1577.782	0.0005	Orbitrap	1	54.3	0.023	3.97	Search stats		GLPGESGAVGPAGPP(OH)GSR	Collagen α2(1)	*T. rex*

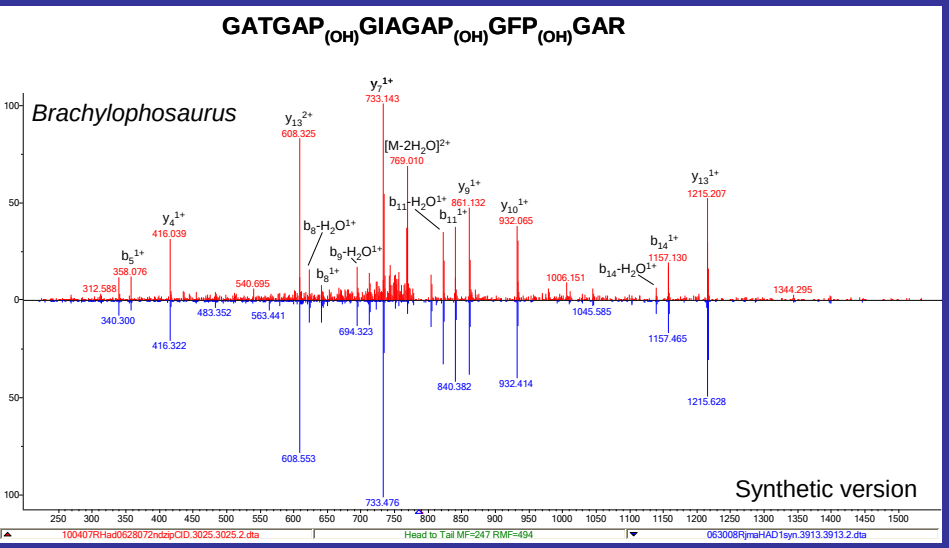
*For two sequences [GLTGPIGPP$_{(OH)}$ GPAGAP$_{(OH)}$ GDKGEAGPSGPPGPTGAR and GATGAP$_{(OH)}$GIAGAP$_{(OH)}$GFP$_{(OH)}$GAR] acquired with the Orbitrap, MS/MS was triggered on the *m/z* ratio representing the ^{13}C stable isotope containing ion rather than the monoisotopic version.

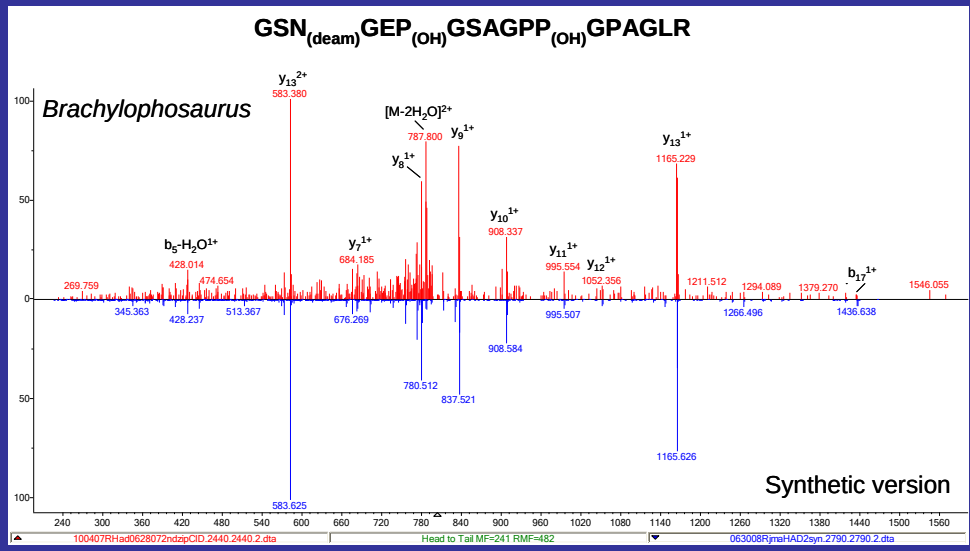
Statistics
(Expectation values<1; FDR 3.04%)

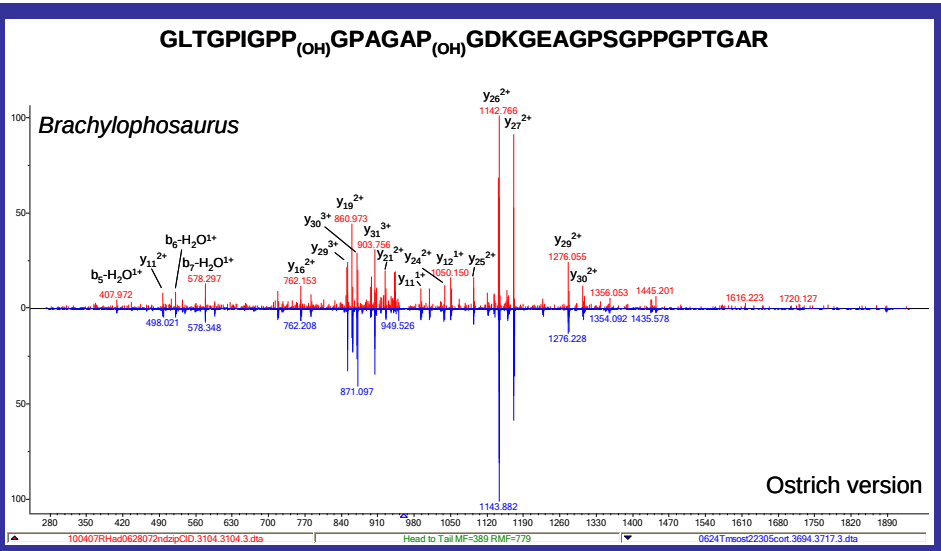
* Sequences confirmed in independent laboratory:
-some sample sent to William Lane, Harvard Univ.

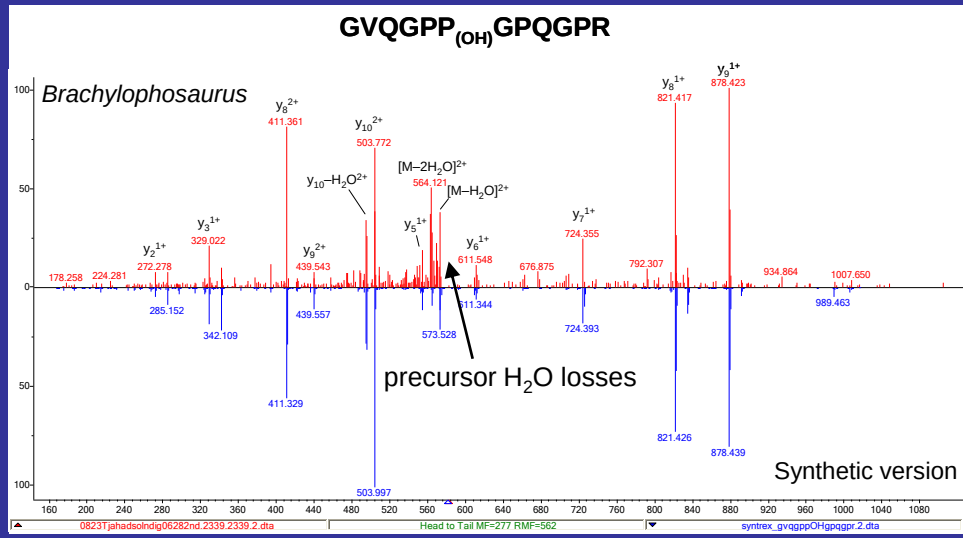
Schweitzer…..Asara, *Science* 2009.

MS Search 2.0 spectral comparison software from NIST





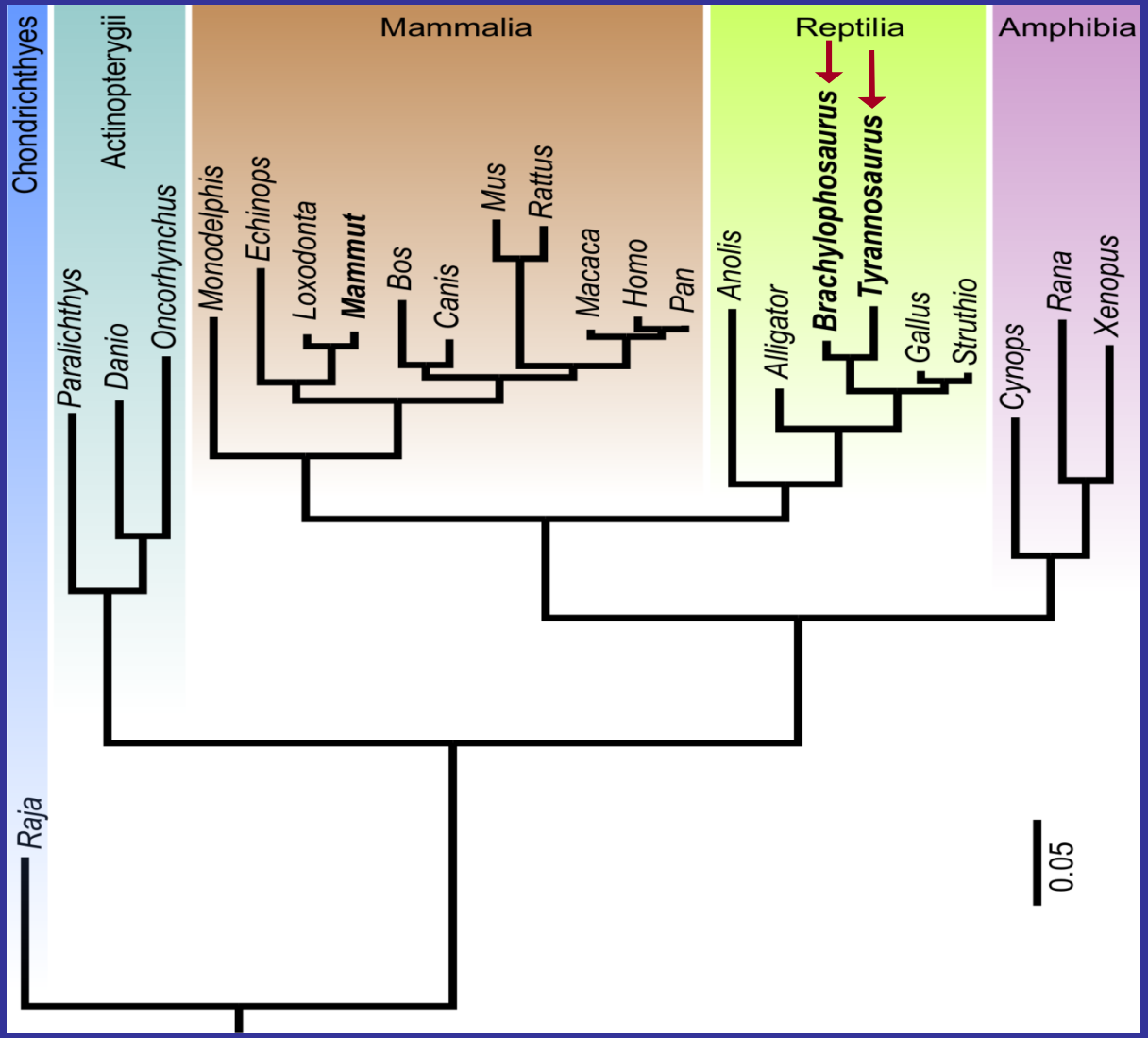




- Validated four lowest scoring Mascot hits vs. Swiss-Prot db
- NIST database of >200,000 spectra including synthetics for comparisons

Schweitzer…..Asara, *Science* 2009.

Phylogenetic Tree of *T. rex* and *Brachylophosaurus* from Collagen Protein Sequences

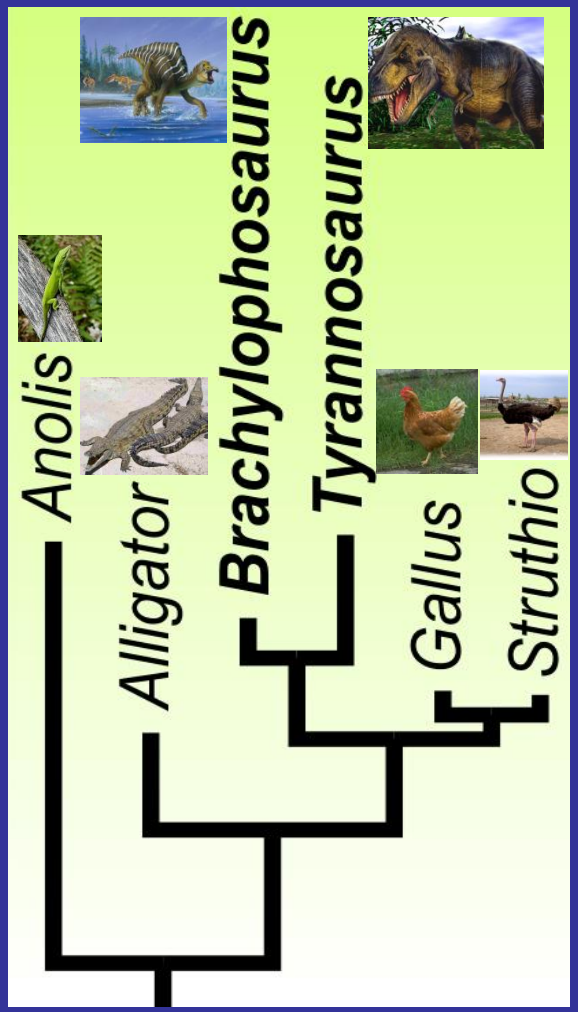


•Bayesian tree validated using additional phylogenetic algorithms (maximum likelihood, parsimony, etc.)

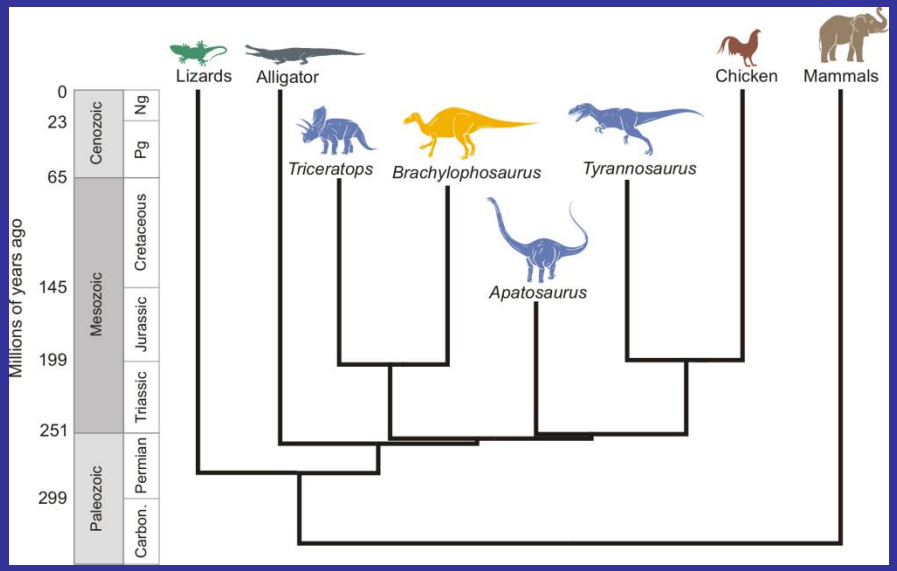
Schweitzer…..Asara, *Science* 2009.

•Phylogenetics analysis groups *Brachylophosaurus* with *T. rex*, very close to birds, ***further from reptiles and lizards***

Collagen Mass Spec sequence prediction

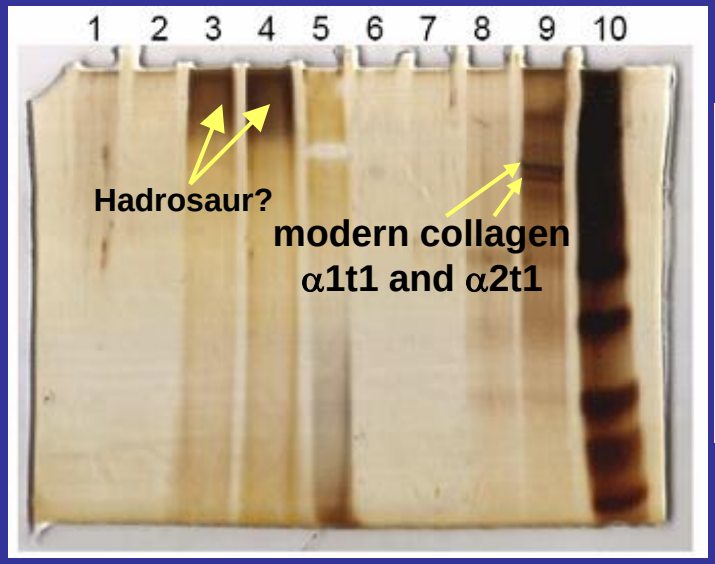


Bone Morphology prediction



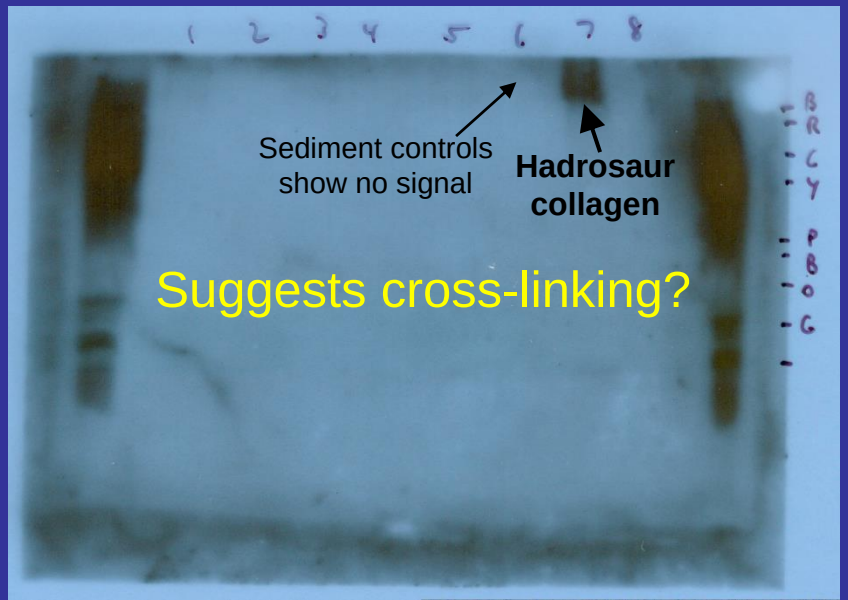
Why and how does protein persist?

SDS-PAGE of 80 million year old Hadrosaur extract shows high MW protein



1. Control extract buffer
2. Sediment
3. Dinosaur extract 1
4. Dinosaur extract 2
5. Dinosaur TCA extract
6. Gel sample buffer
7. Gel sample buffer
8. Ostrich TCA extract
9. Ostrich extract
10. MW markers

Western blot with collagen I alpha 1 antibodies confirm high MW smear is collagen!

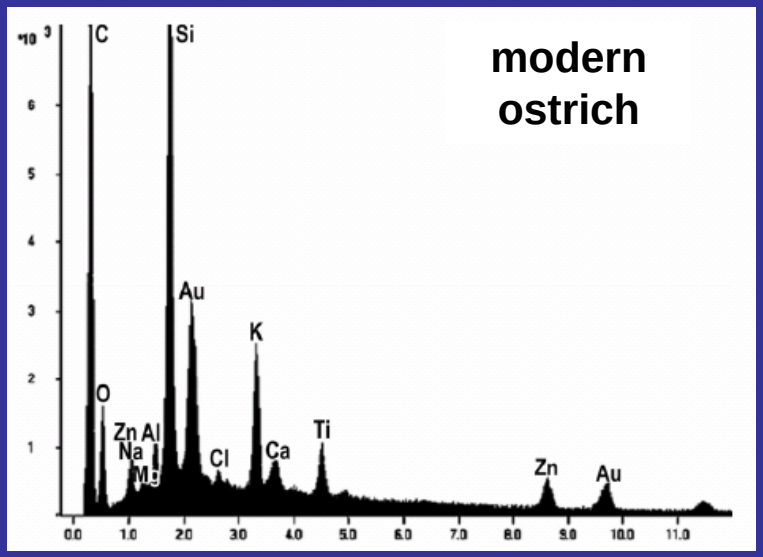


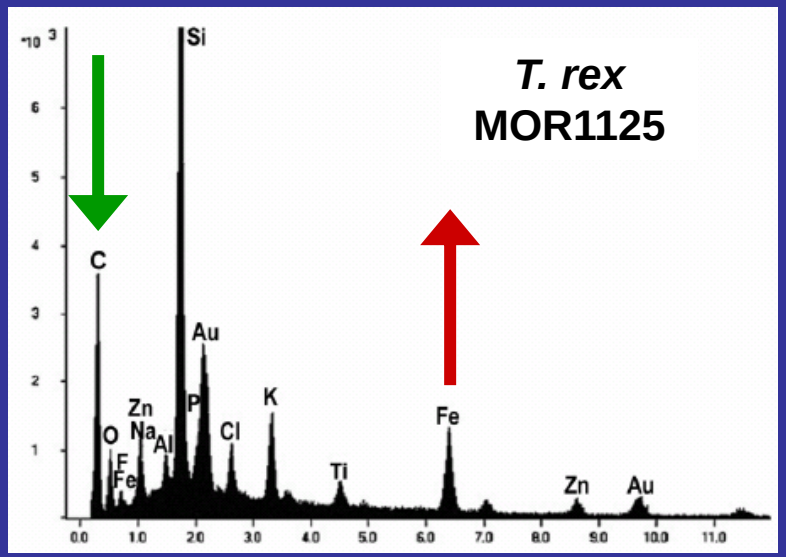
1. Dino 062807 1st extr.
2. Sed 062807 1st extr.
3. Dino 062807 2nd extr.
4. Sed 062807 2nd extr.
5. Dino 070907 1st extr.
6. Sed 070907 1st extr.
7. Dino 070907 2nd extr.
8. Sed 070907 2nd extr.

Schweitzer…..Asara, *Science* 2009.

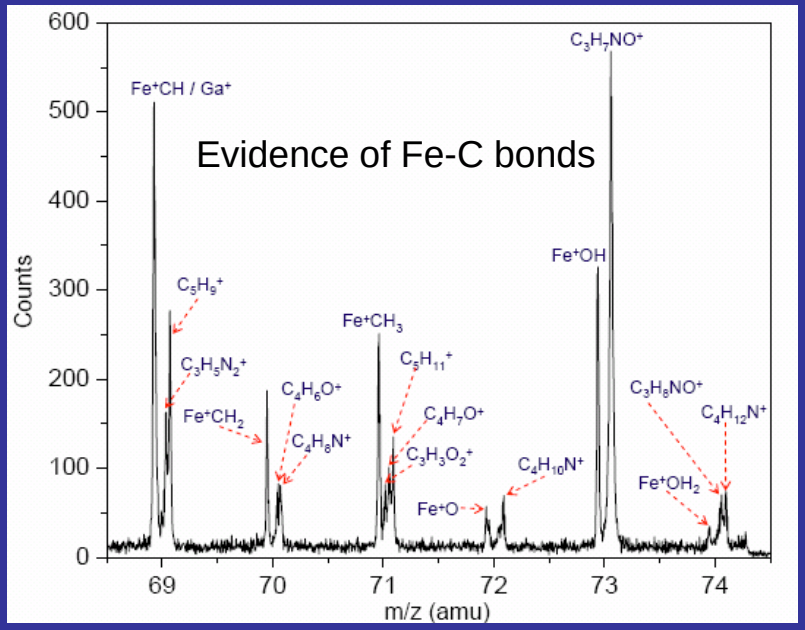
What elements are involved in the protein cross-links?

Elemental analysis





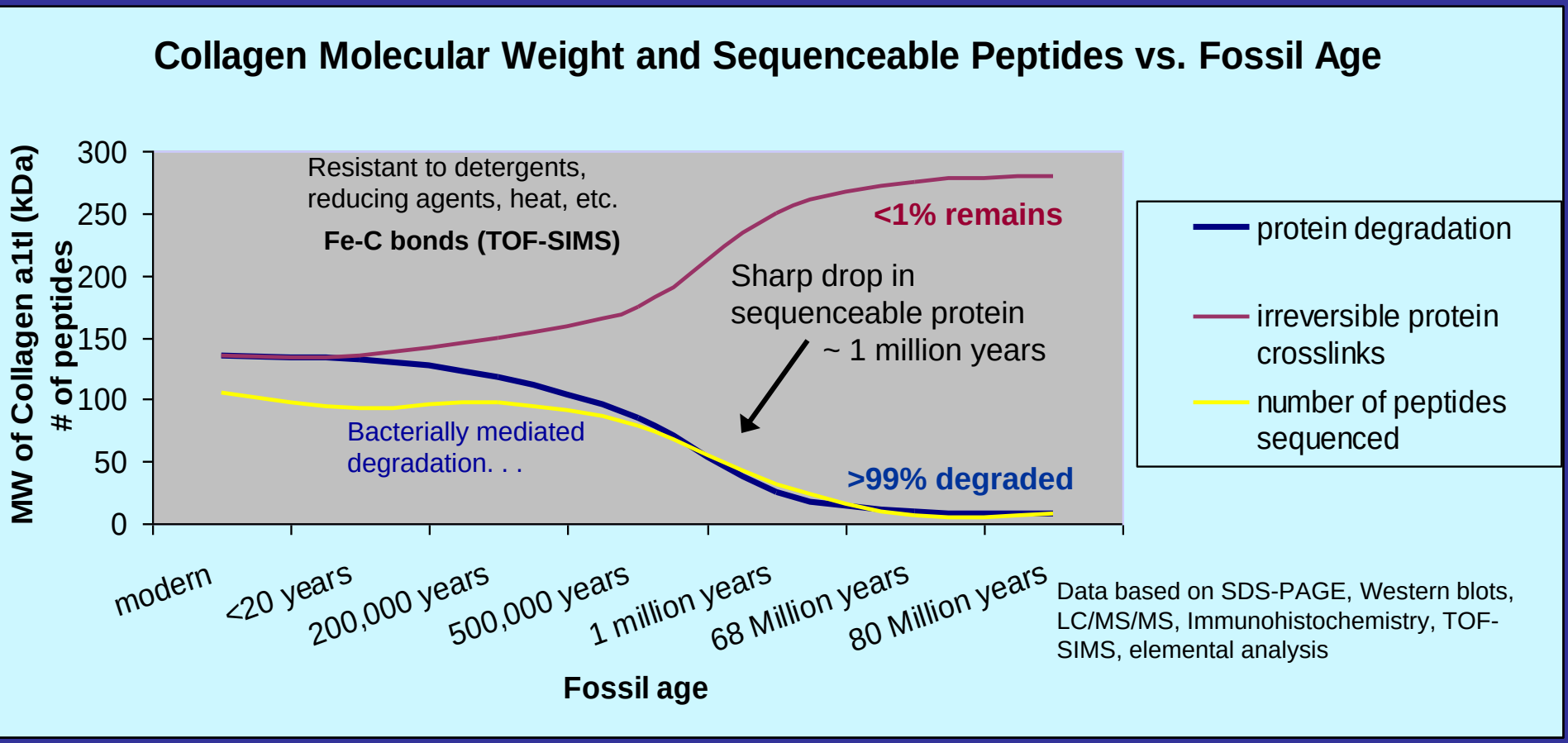
TOF-SIMS: surface analysis



Recep Avci
Montana State Univ.

Schweitzer *et. al*. *Science*, 2005 & 2007.

Model for Protein Preservation: Cross-Linking



- **We are likely sequencing the pieces not involved in irreversible cross-links**

Fossilization Process

Burial in sediment

Protection in hard layer of rock

Fast process for exceptional preservation

Excavation

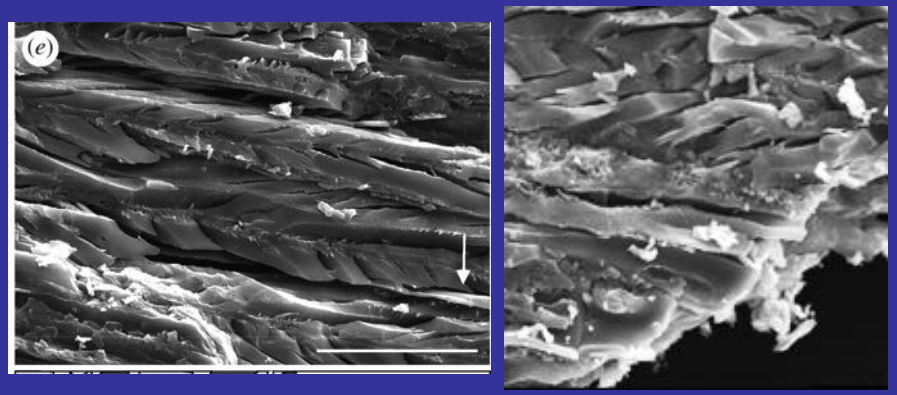
Degradation halts

Fossils were deeply buried in areas *deficient* of the degrading elements: light, water and air

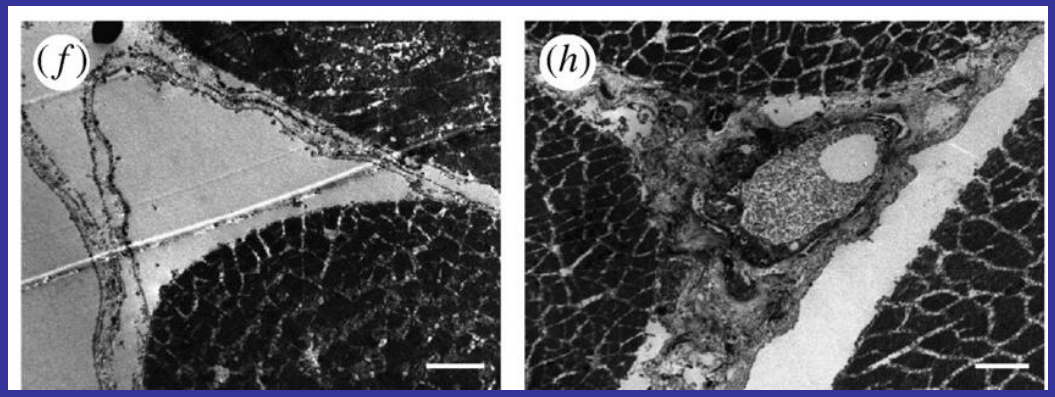
Other recent soft-tissue examples:

Organically preserved musculature in a 18 Myr old fossil salamander *Chelotriton sp.* (MNCN 12555) from Spain

SEM images of muscle fibers



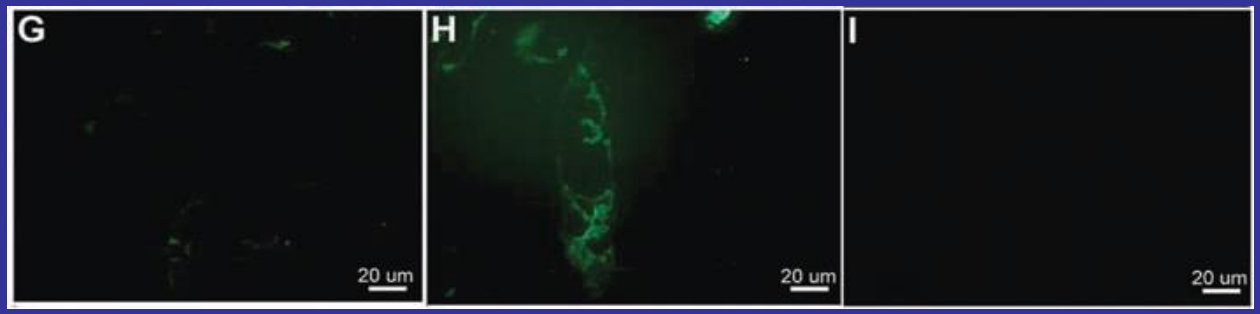
TEM images of transverse sections of skeletal muscle tissue showing circulatory vessels surrounded by endomysium



McNamara *et al*. *Proc. R. Soc. B* 2010;277:423-427

Ongoing studies….

IHC evidence for hemoglobin in 80 million year old *Brachy* bone



We have similar evidence for *T. rex* hemoglobin (to be published…..)

- Get sequence for additional dinosaur proteins for which Mary has immuno data

osteocalcin in bone (Less than 5% of protein content)

elastin in vessels

laminin in vessels

Initial attempts in the Miocene (~1-20 MYA) – mostly marine

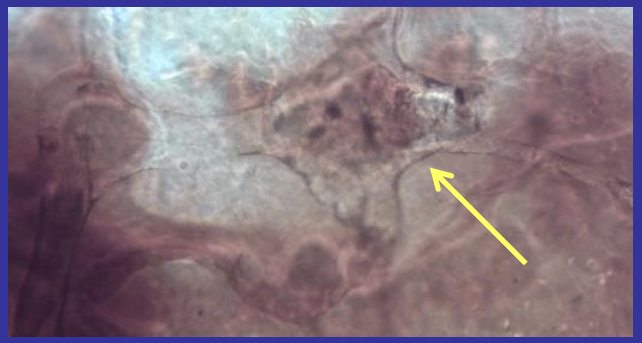
Preservation…preservation….preservation

Crocodile V-3774

Turtle back V-817

Above specimens show no soft-tissue preservation

Manatee (seacow) V-244



- Less than 1% of all fossils have the necessary organic preservation for sequencing – most have been completely mineralized
- Best chance for success may be the Manatee (V-244)

Summary

- It is possible to sequence very *well-preserved* fossil proteins from extinct organisms millions of years old using mass spectrometry
- Proteins may be the only way to molecularly characterize ancient organisms since DNA is not recoverable after ~75K years
- Protein sequences can be used for phylogenetic analyses of ancient organisms such as dinosaur species or *unknown* fossil species

Suggested Reading (besides our papers):

Wired Magazine: Issue 17.07, Origin of Species: *How a T. Rex Femur Sparked a Scientific Smackdown*

-Evan Ratliff

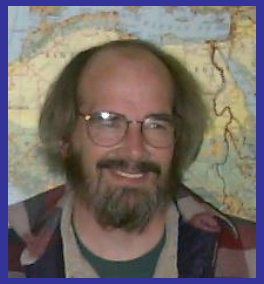
Acknowledgements

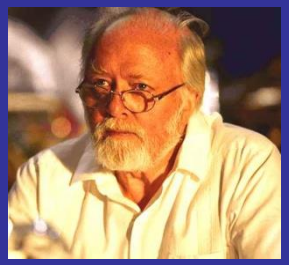
Beth Israel Deaconess Medical Center

Lewis C. Cantley
Lisa M. Freimark
Matthew Phillips
Matthew Vander Heiden
Xuemei Yang

Montana State University

John R. Horner
Christopher L. Hill
Recep Avci





John Hammond (Jurassic Park)

North Carolina State University

Mary H. Schweitzer
Jennifer L. Wittmeyer
Wenxia Zheng

Innovagene Informatics

Shailender Nagpal

Tufts University

Michael Berne

Harvard University

Chris A. Organ
William S. Lane
John Neveu

Funding Sources

National Science Foundation
Paul F. Glenn Foundation
Taplin Funds for Discovery