A Proteomics Jurassic Park:

The isolation of proteins from microorganisms encapsulated in amber from the Oligo-Miocene epoch 30-40 million years ago

Gary B. Smejkal[1,3], George O. Poinar Jr.[2], Feixia Chu[3] and Pier Giorgio Righetti[4]

[1] Harvard Catalyst, Harvard Clinical and Translational Science Center, Laboratory for Innovative Translational Technologies, Boston, MA, USA.
[2] Oregon State University, Department of Zoology, Corvallis, OR, USA.
[3] University of New Hampshire, Hubbard Center for Genome Studies, Durham, NH, USA.
[4] Politecnico di Milano, Department of Chemistry, Materials, and Chemical Engineering, Milan, Italy.

ABSTRACT: Recent reports of peptides isolated from the mineralized skeletal elements and alleged "soft tissues" from *Tyrannosaurus rex* have incited controversy over whether proteins can endure geological time spans. However, soft tissues are not replaced by minerals in amber, which otherwise occurs during lithification. Organisms engulfed in terpenous resins within minutes or seconds were rapidly dehydrated, a prerequisite to preservation, thus retarding any chemical reaction requiring water. Dominican Republic amber from the Oligo-Miocene epoch, 30-40 million years ago, was interrogated for residual proteins. Exclusion of the protein isolates from polyacrylamide gel indicated molecular masses in the multi-million Dalton range, and failure of these aggregates to penetrate these gels proved to be an effective means for further concentrating trace proteins from paleontological samples. Tandem mass spectrometry (LC-MS/MS) of trypsin digests led to the identification of 86 peptides from 20 saprophyte proteins. The most compelling evidence that these peptides are actually of prehistoric origin rather than the result of a contemporary contamination is the extraordinary high degree of crosslinking of these proteins.

1. INTRODUCTION

In 1994, Woodward et al. [1] reported the isolation of DNA fragments from a Late Cretaceous dinosaur bone exhumed from bituminous strata. The following year, the cloning and sequencing of six putative dinosaur DNA fragments derived from a Cretaceous dinosaur egg fossil found in China was later dismissed as the recovered sequences were found to be more closely related to fungi rather than to reptiles or birds [2]. More recently, Asara et al. [3] identified peptides with sequence homology to avian collagen from the mineralized skeletal elements of *Tyrannosaurus rex*. Kaye et al. [4] challenged the finding showing that microbial biofilms formed "endocasts" in which three-dimensional structure was preserved with microscopic detail, but in which the original organic material has been totally replaced by minerals.

The most promising circumstance enabling the preservation of biomolecules over millions of years comes from amber, the fossilized resin of leguminous trees. The unfossilized resins are comprised largely of terpenoids, labdanoids, and phenolics which rapidly dehydrate the included specimen, a prerequisite for preservation, as well as possess anti-bacterial, anti-fungal, and anti-inflammatory properties that intervene with usual decomposition. When the specimen was completely engulfed in the resinous flow, having to occur within seconds, it resulted in unprecedented preservation later observed in amber fossils.

Michael Crichton's novel *Jurassic Park* proposed the recovery of dinosaur DNA from the alimentary tracts of hematophagous insects preserved for millions of years in amber. Though Crichton's work was purely fictional, amber inclusions have shown remarkable preservation of organisms at the tissue and cellular levels, and reptilian blood cells have been identified in partially digested blood meals from parasitic insects encapsulated in Cretaceous amber [5]. Transmission electron microscopy has revealed that subcellular components such as nuclei, endoplasmic reticulum, ribosomes, and mitochondria were still intact in 40 million year old insects preserved in amber [6].

This report describes the isolation of high molecular mass protein aggregates from Oligo-Miocene amber from the Dominican Republic. LC-MS/MS identified several proteins with sequence homology to *Saccharomyces cerevisiae* proteins. While mass spectra does not confirm the source of these peptides, the high degree of protein crosslinking suggests that they are not of contemporary origin.

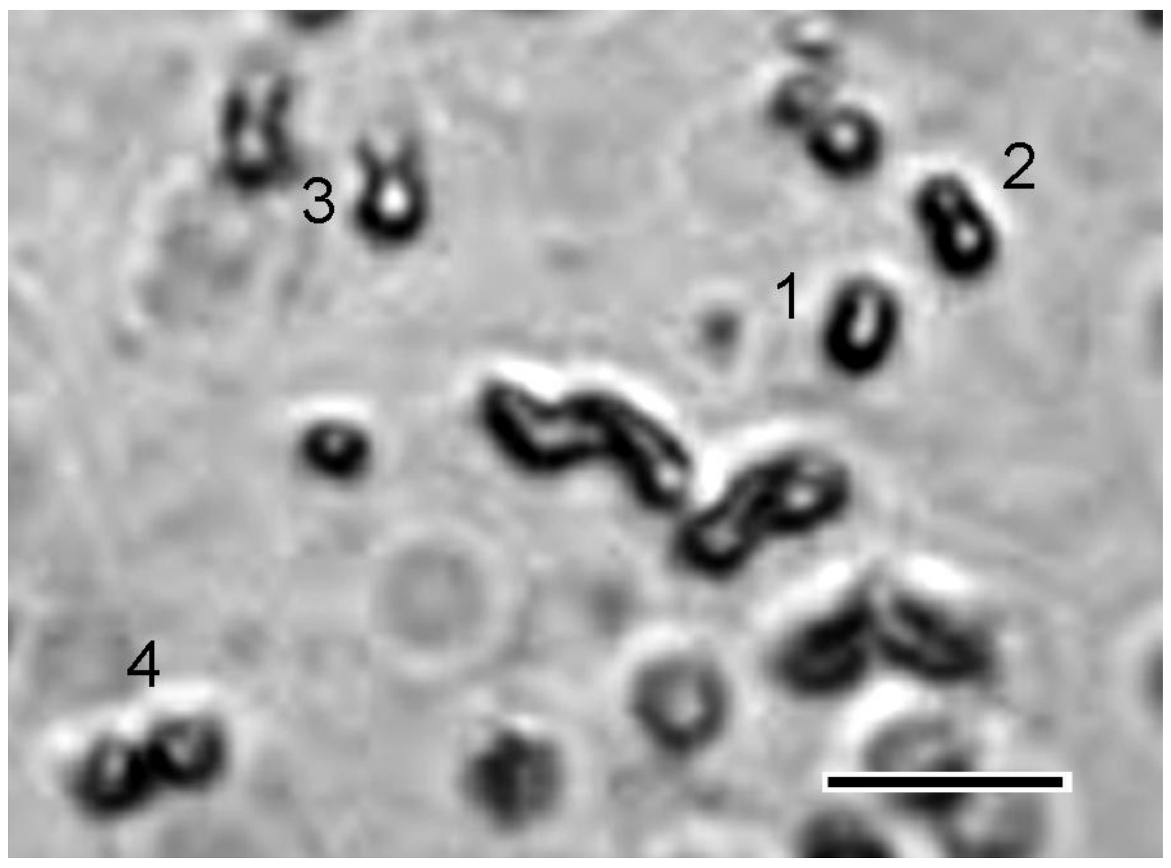

Figure 1. Light microscopy showing budding yeast cells in Dominican Republic amber 20-30 million years old. Stages from single cell **(1)** to where the progeny cell is nearly equal in size to its parent **(4)** and intermediary stages **(2,3)** are shown. Magnification bar equals 6 um.

2 METHODS

2.1 Sample Preparation

All procedures were performed under sterile conditions in a laminar flow hood. Amber pieces containing *Hymenaea protera* leaves were first scrubbed in 5% SDS with a dental brush, then heated to 90° C in 2% SDS and copiously rinsed in water. The amber was rinsed with 100% ethanol just prior to fracture. The amber was fractured and the fragments were ground to a fine triturate using in a sterile Shredder PULSE Tube with serrated metal ram insert (Pressure BioSciences, South Easton, MA). Triturates were extracted in 125 mM Tris-HCl pH 6.8 containing 2% SDS, 5 mM tributylyphosphine, 20 mM AEBSF, 10 mM EDTA and 25 mM phenylacylthiazolium bromide for 100 X 100 seconds at 35,000 psi maximum pressure in a Barocycler NEP 3229 (Pressure BioSciences, South Easton, MA). Samples were filtered in a Ultrafree CL centrifugal filter (Millipore Corporation, Danvers, MA) the filtrates were applied directly to 8-16% polyacrylamide gradient gels (BioRad, Hercules, CA). Gels were stained using the mass spctrophotometry-compatible SilverQuest Silver Stain Kit (Invitrogen, Carlsbad, CA).

Table 1. Peptides identified from 30-40 MYO amber inclusions with sequence homology to *Saccharomyces cerevisiae*

	protein identification	number of peptides
1	enolase 1	11
2	enolase 2	10
3	alcohol dehydrogenase	9
4	glyceraldehyde-3-phosphate dehydrogenase, isozyme 1	8
5	phosphoglycerate kinase	7
6	translation elongation factor EF-1 alpha	7
7	pyruvate kinase	5
8	glyceraldehyde-3-phosphate dehydrogenase, isozyme 3	3
9	heat shock protein YJM789	3
10	fructose 1,6-bisphosphate aldolase	3
11	phosphoglycerate mutase	3
12	ribosomal L12B	2
13	triosephosphate isomerase	2
14	methionine and cysteine synthase (O-acetyl homoserine-O-acetyl serine sulfhydrylase)	2
15	Ssa1p	2
16	mitochondrial aldehyde dehydrogenase	2
17	rubber elongation factor protein	2
18	actin	2
19	40S ribosomal protein S18	2
20	pyruvate decarboxylase	1
	total	86

2.2 LC-MS/MS

Excised gel bands were treated with DTT and iodoacetamide and digested with 100 ng of porcine, side chain-protected trypsin (Promega, Madison, WI). Digestions were carried out at 37ºC for 4 hours. Peptides were extracted with 50 mL of 50% acetonitrile, 2% acetic acid three times and dried down to approximately 10 uL.

Chromatographic separation was carried out on a 75 mm x 15 cm reverse-phase capillary column at a flow rate of 330 nL/min. Solvents (A: 0.1% formic acid in H_2O; B: 0.1% formic acid in acetonitrile) were delivered by a Thermo Surveyor MS Pump through a splitter, running a 5-35% solvent B gradient in 60 minutes. The HPLC eluent was introduced directly to the micro-ion electrospray source of an LTQ Orbitrap XL mass spectrometer (Thermo Fisher Scientific Corporation, Waltham, MA). LC-MS/MS data were acquired in a data dependent acquisition mode, cycling between one precursor ion scan measured in the Orbitrap and six CID scans measured in the LTQ. Activation time was 30 msec. AGC targets were set to 2,000,000 for Orbitrap scans and 10,000 for LTQ MS/MS scans. A dynamic exclusion of 60 seconds duration was selected to prevent repetitive acquisition of high abundance components.

The centroided peak lists of the CID spectra were searched against the National Center for Biotechnology Information (NCBI) protein database using Batch-Tag, a program in the in-house version of the University of California San Francisco Protein Prospector package. Common modifications were considered, including carbamidomethylation of cysteine, N-terminal acetylation, methionine oxidation, and N-terminal glutamine conversion to pyroglutamate. A mass tolerance window of 15 ppm for precursor ions and 0.7 Da for fragment ions was set for database search. The threshold for positive protein identification was at least one peptide with a Protein Prospector peptide score $\geq$ 25 and a peptide expectation value $\leq$ 0.01. Proteins with only two identified peptides (Protein Prospector peptide score $\geq$ 15) were further inspected and confirmed manually.

3. RESULTS AND DISCUSSION

From mass spectra of trypsin digests, 86 peptides with sequence homology to 20 *Saccharomyces cerevisiae* proteins were identified in amber isolates (Table 1). The yeast is likely to be associated with plants and insects embedded in the amber. Experimental procedures for protein extraction, gel electrophoresis and mass spectrometric analysis were tightly controlled with blank samples. The fact that *S. cerevisiae* proteins were identified only in amber samples eliminates the possibility of contemporary contamination. More importantly, all these proteins were identified from a band at the interface of stacking and resolving gel, suggesting an extreme high degree of cross-linking for these proteins during amber formation. Mass spectrometric analysis of another gel band of amber sample did not lead to the identification of any protein. Exclusion of these proteins from 4% polyacrylamide gels indicated molecular masses of several million Daltons, and failure of the aggregates to penetrate these gels proved to be an effective means for concentrating trace proteins from paleontological samples while concomitantly removing interfering substances such as SDS prior to trypsin digestion and LC-MS/MS.

Identifying multiple proteins of *S. cerevisiae* origin was initially surprising. Therefore, peptide sequences of top five proteins were further interrogated to verify the species assignment. For example, enolase 1 (Figure 2) sequences from *S. cerevisiae* and the rubber tree *Hevea brasiliensis* were compared. Although seven of the eleven identified peptides hit conserved regions with high sequence similarity, none of the peptides had identical sequence between these two proteins. Blast NCBI protein database on the other four peptides in divergent regions only returned the saprophyte protein.

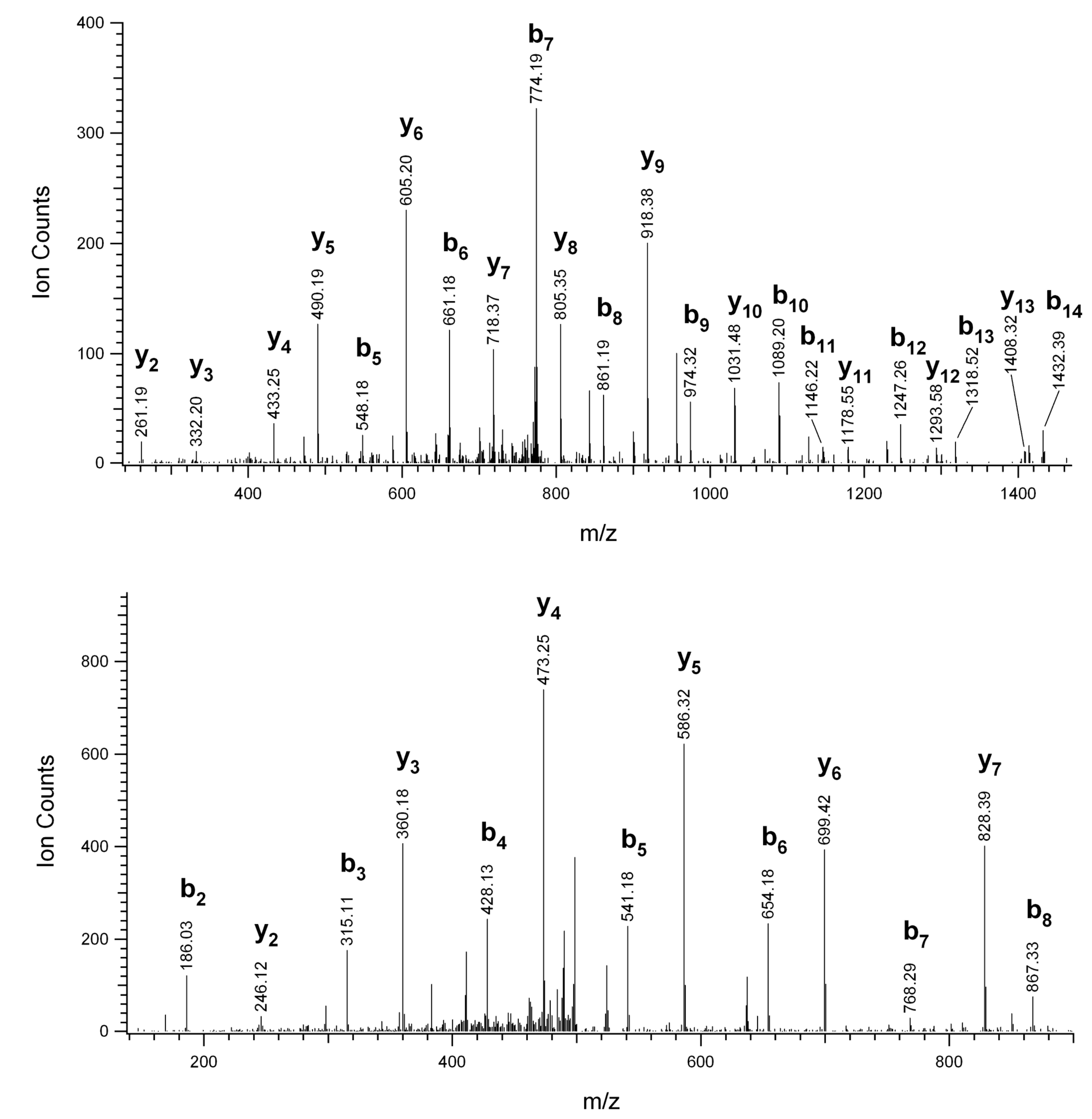

Figure 2. LTQ-Orbitrap mass spectra identifying peptides of sequence homology to *Saccharomyces* enolase I (*top*) and alcohol dehydrogenase (*bottom*) isolated from 30-40 million year old amber.

4. REFERENCES

[1] Woodward S.R. et al. (1994). *Science* 661, 229-232.
[2] Wang H.L. et al. (1997). *Mol. Biol. Evol.,* 14, 589-591.
[3] Asara J.M. et al. (2007). *Science,* 316, 280-285.
[4] Kaye T.G. et al. (2008). *PLoS ONE,* 7, 2808.
[5] Poinar G.O. Jr., Poinar R. (2004). *Vector Borne Zoonotic Dis.,* 4, 281-284.
[6] Poinar G.O. Jr., Hess R. (1982). *Science,* 215, 1241-1242.

THE SIGNIFICANCE OF PALEOPROTEOMICS

The mass extinction of the dinosaurs marking the Cretaceous-Tertiary boundary pales in magnitude compared to other lesser known mass extinction events such as the Permian-Triassic. Five major extinction events, and several smaller ones over the past 540 million years, have resulted in the extinction of 99% of all of the species that ever lived on earth. This is furthered by at least one order of magnitude, since by even the most conservative estimates, less than a tenth of all species presently inhabiting the earth are known, while the vast majority is still undiscovered. Hence, it would appear that the field of biology is based entirely on what has been learned from fewer than 0.1% of all past and present species.

HUPO 9th Annual World Congress, Sydney, Australia, September, 2010. Poster No 543.

PDF available at www.pressurebiosciences.com