

ZooMS: the collagen barcode and fingerprints

Matthew Collins,^a Mike Buckley,^a Helen H. Grundy,^b Jane Thomas-Oates,^a Julie Wilson^a and Nienke van Doorn^a

^aBioArCh, Departments of Biology, Archaeology and Chemistry, the University of York, York YO10 5DD, UK

^bThe DEFRA Food and Environmental Research Agency (FERA), Sand Hutton, York YO41 1LZ, UK

This year the International Barcode of Life initiative (IBoL) plans to begin an ambitious programme to barcode the DNA of more than five million specimens representing at least 500,000 species in five years. Molecular barcodes exploit the fact that molecular sequences offer an independent method to identify a sample. Such molecular barcodes have widespread application in systematics, biodiversity, forensics and even food science. Molecular barcodes tend to be based upon DNA, which with the advent of new technologies offers a fast and efficient means of identification. Proteins too have been used in the past for molecular identification, most commonly exploiting the exquisite specificity of antibodies to discriminate targeted proteins. Recently the idea of using protein mass spectrometry to fingerprint samples has been used to target samples in which processing or decay has destroyed the DNA.

Protein fingerprints

The concept of identification of peptide fingerprints, and indeed amino acid fingerprints before them, has a long history; the work of Klaus Hollemeyer's group (Saarland University, Germany)^{1,2} who have used peptide mass fingerprinting to identify keratin, is notable. Hollemeyer's research group has used mass spectrometry (MS) to identify differences in protein sequences between the pelts of different animals. The hair proteins are cut with enzymes and the masses of the resulting peptides determined. By building up libraries of keratin fingerprints (the dominant protein in hair and feathers), the team is now able

to identify a wide range of feathers and fur, helping in the latter case to enforce a European Union (EU) ban on the use of cat and dog pelts. Other teams have used the same approach to discriminate between species of fish based upon differences in muscle proteins.

In most cases protein fingerprinting methods (see Figure 1) are more expensive than next generation DNA-based screening methods. The value of protein-based identification arises in cases where the DNA is likely to have been destroyed by processing. For example, canning or the processing of pelts and skins leads to fragmentation of DNA, rendering such analyses more difficult.

One process which leads to the fragmentation of molecules is time, and the rates of decay vary between (and even within) molecular groups. Generally speaking, DNA degrades faster than protein, thus in archaeological samples researchers have turned to proteins to aid identification. For example, Hollemeyer and colleagues have successfully applied their methods to identify the origin of the clothing from the famous Copper Age ice mummy Ötzi.^{3,4}

We have used the same logic to develop a tool for bone identification, but working in an archaeological department, archaeological bone was our first target. There were two candidate bone proteins that appeared promising: osteocalcin and collagen. Peggy Ostrom and her team at Michigan State University, USA, were the first group to appreciate the potential of soft ionisation MS of ancient proteins,⁵ and successfully identified osteocalcin in fossil bones from the permafrost.⁶ Mary Schweitzer of North Carolina State

University, USA, and John Asara from Beth Israel Deaconess Medical School, Harvard University, USA, recovered collagen fragments from mammoth,⁷ and then mastodon.⁸ The idea of using protein MS to identify ancient proteins really hit the headlines when Mary and her team claimed to have identified collagen in dinosaur bones,^{8,9} echoing a claim made ten years earlier for the detection of osteocalcin,¹⁰ but this time with the advantage of mass spectra rather than antibody cross-reaction.

We have examined both proteins in archaeological bone,¹¹ and it was evident from this comparative study that collagen was more readily isolated and detected from archaeological bones. The persistence of collagen is well known to archaeologists who routinely use this protein for radiocarbon dating and for stable isotope analysis (in the latter case to obtain dietary and climatic information). The survival mechanism of collagen in archaeological bone is unusual and noteworthy, the stability of the protein being remarkably increased by the formation of a mineral "straight-jacket" which confines the collagen fibrils and prevents them from collapsing.¹² The result is that collagen is a remarkably robust protein when entrapped within bone. It will resist both high temperatures (from processes such as rendering) and long-term burial if the mineral is intact. Indeed long-term burial seems to increase the ease of obtaining a collagen fingerprint as over time other proteins such as haemoglobin and osteocalcin are lost more rapidly, resulting in the selective enrichment of collagen in old bone.

© 2010 Thermo Fisher Scientific Inc. All rights reserved. All trademarks are the property of Thermo Fisher Scientific Inc. and its subsidiaries.

The standard in FT-IR Spectroscopy, with all the options.

When you need a rugged, high performance FT-IR spectrometer, you want the Thermo Scientific Nicolet™ 6700. Or for the most advanced experiments, look to the Nicolet 8700 FT-IR spectrometer.

For years the name Nicolet has been synonymous with spectrometer innovation. We introduced you to Smart™ Accessories, OMNIC™ software, utilities for nearly any application known to FT-IR, and the highest performing optics available. Today the Thermo Scientific brand represents the gold standard for reliable, repeatable, results-driven operation.

When new materials and new challenges cross your lab, be certain that your Thermo Scientific FT-IR spectrometer is ready to deliver solid answers.

Visit www.thermo.com/getgold today.



**Thermo Scientific Nicolet 6700
FT-IR Spectrometer**

FASTLINK / CIRCLE 003 FOR FURTHER INFORMATION

Moving science forward

Thermo
SCIENTIFIC

Part of Thermo Fisher Scientific

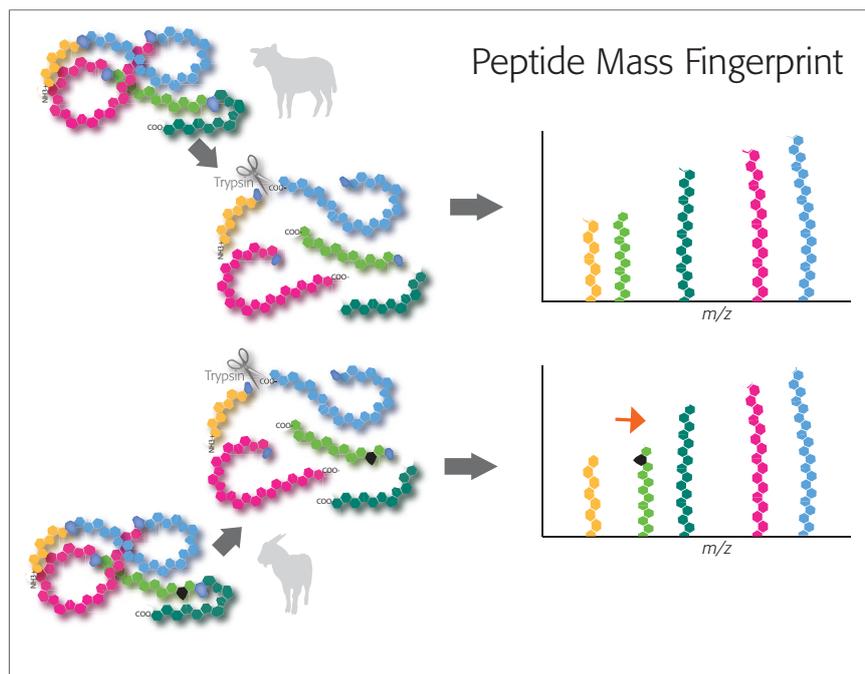


Figure 1. The principle of peptide mass fingerprints. Species can be discriminated by the difference in mass of that peptide caused by the substitution of an amino acid, which changes the mass of the peptide.

... and he will separate them one from another, as the sheep from the goats

It is possible to identify bones by comparing collagen peptide fingerprints with the fingerprints from known samples. In the case of collagen there are two particular problems, first, very few collagen sequences are known, and second, collagen has extensive post-translational modification, notably as a result of the process of hydroxylation of the abundant proline residues.

We have built up a library of collagen sequences using a combination of molecular mining of EST (Expressed Sequence Tags) and other databases, coupled with *de novo* sequencing, so that it is now possible to identify a wide range of animal species using peptide mass fingerprints (Figure 2, see also Reference 13). The method has been successfully applied to modern and archaeological bone samples, and is now being tested with bone fragments recovered from animal feed (meat and bone meal). This latter kind of analysis reveals the power of a mass spectrometric approach, exploiting the speed

and accuracy of matrix-assisted laser desorption/ionisation mass spectrometry (MALDI-MS) to rapidly identify individual bone particles, in order to ensure that banned ruminant tissue is not present in animal feed. Despite the obvious potential of this approach our focus is archaeological. On archaeological sites many bones are variously chewed, cooked, trampled or crushed. This processing removes many of the distinguishing features of the bone and consequently a large fraction is no longer identifiable. We propose the use of peptide fingerprinting to solve this problem, an approach we term ZooMS, short for Zooarchaeology by Mass Spectrometry. We have already used the method to discriminate sheep from goats,¹⁴ a problem in archaeology (of biblical proportion) that has led to the widespread use of the term "ovicaprine" to describe the very many bones that resist identification. The different husbanding practices and exploitation practices of the two groups are hidden if the two species cannot be separated.

Collagen is a slowly evolving protein consisting of three chains wound together as a triple helix. In mammals and most other vertebrate groups two chains are

from the same COL1A1 gene, with a third, more rapidly evolving, COL1A2, chain. Careful analysis of the difference between sheep and goat collagen reveals that there are two differences in the sequence lying close together on the COL1A2 chain, the double mutation arising at the base of the genus *Capra*.¹⁴ The slow evolution of the collagen chains means that whilst ZooMS can sort the sheep from the goats, this means *all* goats, it cannot tell the difference between a domestic goat and a wild ibex.

Collagen, the barcode of death?

The slow speed of evolution of the collagen chains means that ZooMS will have a resolution that approximates to genus (although clearly this will depend upon taxonomic group). However, the large size of the protein and the slow rate of evolution means that it has sufficient variation to both: (i) be able to discriminate between genera across the animal kingdom using the same protein and the same extraction methods, and (ii) is sufficiently similar to be able to map differences across widely-dispersed groups. The fact that collagen is selectively preserved in old bone (and teeth, ivory and antler) means that it can be used as a molecular barcode long after DNA barcodes have been damaged, fragmented and have "melted" away. Collagen can be recovered from most archaeological bone and, if the work of Mary Schweitzer and her colleagues bears fruit, from fossils of far greater antiquity. This means that collagen is an excellent barcode for these animal tissues, both from processed tissue, such as meat and bone meal, but also traded goods, such as antler and ivory as well as from archaeological bones and the fossils of many animals that suffered a wave of extinction as modern humans migrated out of Africa and across the globe approximately 70,000 years ago.

The particular characteristics of mass spectrometry, the ability to detect original molecular sequences without amplification (and attendant risks of contamination) and the speed and simplicity of analysis, has the potential to revolutionise the identification of animals' tissues, even from tiny fragments, far back into

Maximum Performance



... at Maximum Speed

Analytical Power and Speed for:

- Fast chromatography applications
- Trace analysis studies
- Small molecule ID
- Metabolomics
- Quantitative proteomics

**** >40.000 FWHM resolution ** <1 ppm mass accuracy ** 20 spectra/second ****

It's time to embrace a new concept of what is possible with a mass spectrometer. The maXis Ultra High Resolution (UHR) - TOF is the only tandem mass spectrometer able to provide maximum information at the very highest speeds delivered by latest Ultra Performance Liquid Chromatography. No other mass spectrometer is better equipped to acquire definitive data for small molecule identification, proteomics and metabolomics applications.

Contact us now and get involved with the maXis today at www.bdal.com/maxis

think forward

UHR-TOF MS

FASTLINK / CIRCLE 004 FOR FURTHER INFORMATION

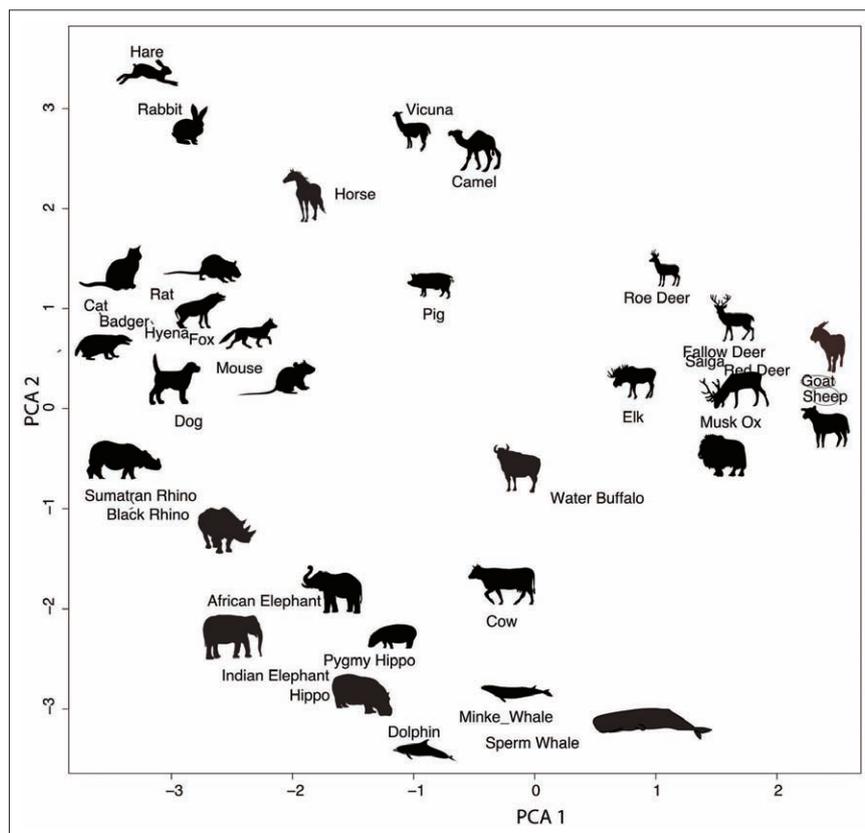


Figure 2. Principal components analysis (PC scores-based plot) of bone collagen peptide masses determined by MALDI-time-of-flight-MS reveals good discrimination between mammals, despite the fact that Type I collagen is a relatively well conserved protein.

time and across the vertebrate kingdom. We would argue that just as DNA can be considered a barcode of life, collagen has the potential to be the barcode for the communities of the dead.

References

1. K. Hollemeyer, W. Altmeyer and E. Heinzle, "Identification of furs of domestic dog, raccoon dog, rabbit and domestic cat by hair analysis using MALDI-ToF mass spectrometry", *Spectrosc. Europe* **19**(2), 8–15 (2007).
2. K. Hollemeyer, W. Altmeyer and E. Heinzle, "Identification and quantification of feathers, down, and hair of avian and mammalian origin using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry", *Anal. Chem.* **74**(23), 5960–5968 (2002).
3. K. Hollemeyer, W. Altmeyer, E. Heinzle and C. Pitra, "Species identification of Oetzi's clothing with matrix-

assisted laser desorption/ionization time-of-flight mass spectrometry based on peptide pattern similarities of hair digests", *Rapid Commun. Mass Spectrom.* **22**(18), 2751–2767 (2008).

4. K. Hollmeyer, W. Altmeyer, E. Heinzle and C. Pitra, "Species origin identification of Oetzi's clothing by MALDI-ToF mass spectrometry using tryptic hair digests", *Spectrosc. Europe* **21**(2), 7–11 (2009).
5. P.H. Ostrom, M. Schall, H. Gandhi, T.L. Shen, P.V. Hauschka, J.R. Strahler and D.A. Gage, "New strategies for characterizing ancient proteins using matrix-assisted laser desorption ionization mass spectrometry", *Geochim. Cosmochim. Acta* **64**(6), 1043–1050 (2000).
6. C.M. Nielsen-Marsh, P.H. Ostrom, H. Gandhi, B. Shapiro, A. Cooper, P.V. Hauschka and M.J. Collins, "Exceptional preservation of bison bones >55 ka as demonstrated by protein and DNA

sequences", *Geology* **30**(12), 1099–1102 (2002).

7. M.H. Schweitzer, C.L. Hill, J.M. Asara, W.S. Lane and S.H. Pincus, "Identification of immunoreactive material in mammoth fossils", *J. Molecular Evolution* **V55**(6), 696–705 (2002).
8. J. Asara, M. Schweitzer, L. Freimark, M. Phillips and L. Cantley, "Protein sequences from mastodon and *Tyrannosaurus rex* revealed by mass spectrometry", *Science* **316**(5822), 280 (2007).
9. M.H. Schweitzer, W. Zheng, C.L. Organ, R. Avci, Z. Suo, L.M. Freimark, V.S. Lebleu, M.B. Duncan, M.G. Vander Heiden, J.M. Neveu, W.S. Lane, J.S. Cottrell, J.R. Horner, L.C. Cantley, R. Kalluri and J.M. Asara, "Biomolecular characterization and protein sequences of the Campanian hadrosaur *B. canadensis*", *Science* **324**(5927), 626–631 (2009).
10. G. Muzer, P. Sandberg, M.H.J. Knapen, C. Vermeer, M.J. Collins and P. Westbroek, "Preservation of the bone protein osteocalcin in dinosaurs", *Geology* **20**, 871–874 (1992).
11. M. Buckley, C. Anderung, K. Penkman, B.J. Raney, A. Götherström, J. Thomas-Oates and M.J. Collins, "Comparing the survival of osteocalcin and mtDNA in archaeological bone from four European sites", *J. Archaeolog. Sci.* **35**(6), 1756–1764 (2008).
12. A.D. Covington, L. Song, O. Suparno, H. Koon and M.J. Collins, "Link-lock: an explanation of the chemical stabilisation of collagen", *J. Soc. Leather Technol. Chem.* **92**(1), 1 (2008).
13. M. Buckley, M. Collins, J. Thomas-Oates and J.C. Wilson, "Species identification by analysis of bone collagen using matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry" *Rapid Commun. Mass Spectrom.* **23**(23), 3843–3854 (2009).
14. M. Buckley, S. Whitcher Kansa, S. Howard, S. Campbell, J. Thomas-Oates and M. Collins, "Distinguishing between archaeological sheep and goat bones using a single collagen peptide", *J. Archaeolog. Sci.* **37**, 13–20 (2010).