

# Collagen survival and its use for species identification in Holocene-lower Pleistocene bone fragments from British archaeological and paleontological sites

Mike Buckley,<sup>1,2</sup> Matthew James Collins<sup>2</sup>

<sup>1</sup>Manchester Interdisciplinary Biocentre, Faculty of Life Sciences, Manchester;

<sup>2</sup>BioArCh, Departments of Biology, Archaeology and Chemistry, University of York, Heslington, UK

## Abstract

Proteins have long been known to persist in Quaternary bone fossils and are often targeted as a source of carbon used in radiocarbon dating and stable isotope analyses for determining provenance and obtaining dietary information. We have previously reported a technique using the dominant structural protein collagen (type I) as a source of genetic information for species identification in modern and relatively young (Holocene) archaeological samples. We report a systematic investigation of amino acid composition and collagen peptide mass fingerprints (PMF), for a range of samples dating back approximately 1.5 million years. Extrapolation from high temperature experimental decomposition rates predict that at a constant 10°C (the approximate mean annual air temperature in Britain today) it will take between 0.2 and 0.7 Ma for levels of collagen to fall to 1% of their original concentration in an optimal burial environment. Even when the glacial intervals of the British Quaternary are factored into the temperature calculations, the more conservative of these two estimates extends the range for collagen sequencing to the Lower Pleistocene as confirmed by the presence of collagen peptides in bones from the Weybourne Crag (~1.5 Ma). Collagen fingerprinting can extend the range of identifiable taxa present at sites with large assemblages of fragmentary bone material such as that encountered at the ~900 Ka site at Happisburgh (Norfolk, UK) recently identified as showing signs of the earliest humans in Britain.

## Introduction

Species identification of animal fossils contributes to the understanding of how humans interacted with and managed animals in the

more recent past, as well as how changing environmental conditions affected animal populations in the more distant past. Much research has focused on the British Quaternary period because of the extreme variations in climate history coupled with the recurring insularization that resulted in relatively unique combinations of animals that collectively can be diagnostic of a particular period (Mammal Assemblage Zones).<sup>1</sup> However, many archaeological and palaeontological assemblages have high numbers of fragmentary bone finds that vastly outnumber those that remain intact. The fragmentation of bone via attrition often results in a complete loss of morphologically-diagnostic criteria used for species identification, which limits the potential to confer archaeological and/or environmental information.

We have previously applied the concept of a molecular fingerprint to the identification of modern bone fragments using peptide mass fingerprints (PMFs) of collagen (ZooMS) (Supplementary Table 1).<sup>2</sup> Collagen (I) by its very nature, as a highly insoluble structural protein, is a robust protein that is particularly resistant to decay. Support for its longevity in the optimal burial environment has come in the form of mass spectrometric sequencing analyses from Middle Pleistocene (~780-125 Ka) samples<sup>3,4</sup> and suggested survival from stable isotope results of Lower Pleistocene bones.<sup>5</sup> Despite early interest in the analysis of collagen from fossil bone<sup>6</sup> there have been no systematic surveys of the persistence of collagen beyond the radiocarbon limit (~50 Ka).

In this paper we explore the potential of using collagen fingerprinting to identify archaeological and palaeontological samples throughout the Holocene and Pleistocene dating back to ~1.5 million years. In order to understand the potential of this method for ancient (sub-fossil) samples, we compare rates of collagen loss with predictions derived from laboratory experiments to assess the quality and thus capability of the collagen analyses to be used in identifying bone fragments from British sites dating back to the Lower Pleistocene. We test this model using samples from a range of sites that span 1.5 Ma. These samples include examples of how the technique could serve as a means of testing uncertain morphological identifications and as such have the potential to significantly alter interpretation particular and assemblages (e.g., Wetwang Slack, East Yorkshire) or as a way of identifying new species at a site where material is too fragmentary to offer morphological information at any meaningful taxonomic level (e.g., at Happisburgh, Norfolk, which currently has the earliest evidence of humans in Britain).

Correspondence: Mike Buckley, Manchester Interdisciplinary Biocentre, Faculty of Life Sciences, 131 Princess Street, Manchester, M1 7DN, UK. E-mail: m.buckley@manchester.ac.uk

Key words: collagen, Pleistocene, diagenesis, MALDI-MS, ZooMS.

Acknowledgements: the authors would like to thank Mandy Jay, the Hull & East Riding Museum, Andy Curren, Adrian Lister, Andrew Chamberlain, Terry O'Connor, Nigel Larkin, the Yorkshire Museum, the University of Nottingham Archaeology Museum for access to the samples used in this study, Colin Smith for access to data on heated cattle bone and Sarah Howard, Colin Smith, Kirsty Penkman and Oliver Craig for commenting on early versions of the manuscript. We also acknowledge the support of the Centre of Excellence in Mass Spectrometry (University of York) and the North East Amino Acid Racemization Laboratory (BioArCh, University of York) for access to instrumentation. This work was supported by the Natural Environment Research Council [NER/S/J/2004/13017 and NE/G000204/1].

Received for publication: 29 June 2011.

Accepted for publication: 14 September 2011.

This work is licensed under a Creative Commons Attribution 3.0 License (by-nc 3.0).

©Copyright M. Buckley and M.J. Collins, 2012  
Licensee PAGEpress, Italy  
Antiqua 2011; 1:e1  
doi:10.4081/antiqua.2011.e1

## Materials and Methods

To widen the range of species in our database of collagen fingerprints, modern bone samples of arctic fox (*Alopex lagopus*), brown bear (*Ursus arctos*), lion (*Panthera leo*), reindeer (*Rangifer tarandus*), horse (*Equus caballus*), grey seal (*Halichoerus grypus*), and narwhal (*Monodon monoceros*), were obtained as standards, whilst some species not previously analysed for collagen fingerprinting, including lynx (*Lynx sp.*), bison (*Bison sp.*), beluga whale (*Delphinapterus leucas*), harp seal (*Pagophilus groenlandicus*) and human (*Homo sp.*) were obtained from archaeological samples. In order to cover a wide temporal range, samples from many different archaeological and palaeontological sites and burial environments were obtained, ranging from the Iron Age (Wetwang Slack, East Yorkshire; ~200 B.C.), Neolithic (Carsington Pasture Cave and Totty Pot), Late-Glacial (Etches Cave & Fox Hole Cave), Middle Devensian (Uphill Cave), Early Devensian (Steeley Wood Cave and Stump Cross Caverns), Ipswichian (Kirkdale Cave), the Cromerian (Kessingland Forest Bed and

West Runton) Forest Bed and the underlying Weybourne Crag deposits (~1.5 Ma) as well as several samples of uncertain age dredged from the North Sea.

Hydrochloric acid (HCl), ammonium bicarbonate, and acetonitrile (ACN) were obtained from BDH (UK). Sequencing grade trypsin was purchased from Promega (UK). Bacterial collagenase, trifluoroacetic acid (TFA), mass spectrometric standards and  $\alpha$ -cyano-4-hydroxycinnamic acid were purchased from Sigma-Aldrich (UK). C18 solid phase extraction (SPE) pipette tips were purchased from OMIX (UK) and C18 SPE 100 mg cartridges from Amersham Biosciences (UK). The six MALDI-ToF-MS calibration peptides were purchased from Applied Biosystems.

### Modeling collagen degradation

To simulate the thermal degradation of type I collagen, samples of modern cattle (*Bos taurus*) and human (*Homo sapiens*) bone pieces were heated in sealed Pyrex glass vials under wet conditions (10 mL distilled water 1 g bone) at 90°C for a maximum duration of 32 days.<sup>7</sup> Sub-samples, including 5 mL of the supernatant were taken after 1, 2, 4, 8, 12, 16, 20, 26 and 32 days of incubation. After heating, the vials were cooled, decanted and lyophilised for subsequent amino acid analysis steps.

### Collagen fingerprint analysis (ZooMS)

Collagen fingerprints were obtained following modified methods to those described elsewhere.<sup>2</sup> Briefly, samples of bone were powdered, demineralized (1 mL 0.6 M HCl per 50 mg sample, 4 h, 4°C) and centrifuged (13,000 x g). The acid-insoluble residue was then heated at 70°C in 50 mM ammonium bicarbonate for 3 h and following centrifugation (15 min, 13,000 x g) the supernatant was digested with trypsin (37°C for 18 h). Following digestion, the samples were centrifuged (10 min; 13,000 x g), the supernatant acidified (0.1% TFA) and loaded on to activated C18 in pipette tip (activated with 2 x 100  $\mu$ L 50% in 0.1% TFA followed by 2 x 100  $\mu$ L 0.1% TFA). Peptide fractions were eluted following several wash steps (2 x 0.1% TFA) using a stepped gradient (10% followed by 50% ACN in 0.1% aqueous TFA). Eluted fractions (1  $\mu$ L) were spotted onto a Bruker ultraflex III target plate, mixed together with 1  $\mu$ L of  $\alpha$ -cyano-4-hydroxycinnamic acid matrix solution (1% in ACN/H<sub>2</sub>O 1:1 v:v) and allowed to dry. Each spot was analysed by MALDI-MS using a Bruker ultraflex III MALDI TOF/TOF mass spectrometer equipped with a Nd:YAG smart beam laser.

### Collagen telopeptide analysis

The  $\alpha$ 2 (I) chain carboxytelopeptide was isolated following<sup>8</sup> in order to assess the qual-

ity of the surviving protein following the hypothesis that they would likely be the first to show signs of degradation.<sup>9</sup> In brief, ~1 mg demineralised collagen (as above) was digested with bacterial collagenase for 5 h and isolated using gravity-flow solid phase extraction (SPE) C18 columns eluted (following two wash steps of 0.1% TFA and one of 20% ACN in 0.1% TFA) with 26% ACN in 0.1% aqueous TFA. The samples were then concentrated by evaporation and rehydrated with 0.1  $\mu$ L 0.1% TFA and 0.5  $\mu$ L spotted (wet droplet method) with 0.5  $\mu$ L of  $\alpha$ -cyano-4-hydroxycinnamic acid matrix solution (1% in ACN/H<sub>2</sub>O 1:1 v:v), allowed to dry and then analysed by MALDI-MS using an Applied Biosystems 4700 Proteomics Analyzer.

### MALDI-MS analysis

The fractions of each collagen digest were analysed by Matrix Assisted Laser Desorption Ionization – Time of Flight – Mass Spectrometry (MALDI-TOF-MS) in reflectron mode. Calibrated MS spectra were acquired over a mass range of  $m/z$  800-4000. Monoisotopic masses were obtained using a S/N threshold of 2. *De novo* sequencing of selected collagen peptides was carried out by manually interpreting the spectra obtained on collision-induced dissociation product ion (MS/MS) of selected peptides (e.g., Supplementary Figure 1). Modern collagens of species not previously analysed were evaluated for additional peptide markers and added to establish species markers.<sup>2</sup> Analyses of all ancient samples were matched against the combined set of species markers.

### Amino acid analysis

Amino acid composition analyses were carried out following.<sup>10</sup> In brief, ~1 mg bone powder was treated with 100  $\mu$ L of 7 M HCl under N<sub>2</sub> at 110°C for 18 hours. Samples were then evaporated and rehydrated with 500  $\mu$ L 0.01 mM L-homo-arginine solution. An aliquot was then injected and mixed online with 2.2  $\mu$ L derivatising reagent (260 mM n-iso-butyl L-cysteine and 170 mM o-phthalaldehyde in 1 M potassium borate buffer adjusted to pH 10.4) and separated on a C18 HyperSil BDS column (5 x 250 mm) at 25°C using a gradient elution of sodium acetate buffer (23 mM sodium acetate trihydrate, 1.5 mM sodium azide, 1.3 mM ethyldiamine tetra-acetic acid, pH 6), methanol and acetonitrile.

## Results

### Taxonomic resolution of ZooMS

We have previously reported the use of PMFs to discriminate between bone collagens typically using only seven of the >100 tryptic colla-

gen (I) peptides usually observed, selected because they display useful sequence variation and are routinely detected in MALDI-TOF-MS.<sup>2</sup>

### Carnivora

Within the taxonomic order Carnivora we observed sequence variation being between morphologically-similar arctic (*Alopex lagopus*) and red fox (*Vulpes vulpes*), which are often difficult to discriminate in degraded palaeontological assemblages but can infer different ecological niches/environments. The estimated divergence of these two taxa is ~5 Ma,<sup>11</sup> yet they are sometimes placed in the same genus (*Vulpes*).<sup>12</sup> Other closely related carnivores investigated are the feliformes, including *Panthera*, *Lynx* and *Felis*. All three groups diverged ~6-7 Ma<sup>13</sup> and can be differentiated using analogous peptide markers to those we have previously described<sup>2</sup> but two seal species (Harp seal (*Pagophilus groenlandicus*) and grey seal (*Halichoerus grypus*)), which also diverged approximately ~6-7 Ma<sup>13</sup> could not be differentiated.

### Artiodactyla

Within the artiodactyls, we were unable to discriminate using the seven peptide markers between *Bos* and *Bison* where estimated divergence is ~1-4.5 Ma.<sup>14</sup> It was possible to discriminate the three tribes of Cervidae where the approximate divergence times of the Capreolini (roe deer, *Capreolus capreolus*) and Rangiferini (reindeer, *Rangifer tarandus*) lie close to that of the divergence of these two tribes from the sub-family Cervini (red deer, *Cervus elaphus*; fallow deer, *Dama dama*) at ~7 Ma.<sup>15</sup>

### Cetacea

Of the six cetacean species analysed, the baleen whale (Minke whale), which diverged from toothed whales ~35-38 Ma,<sup>16</sup> possessed several collagen peptide markers not seen in any previously studied species. Of the three toothed (Odontoceti) whale species, the sperm whale, which diverged from the remaining toothed whales at ~35 Ma, displayed the most distinct peptide sequences. The two dolphin species Risso's and Euphrosyne separated from the remaining toothed whale lineage at ~18 Ma and also shared diagnostic dolphin peptides as previously reported.<sup>2</sup> However, we were unable to discriminate between these two dolphins (divergence of ~6 Ma).<sup>16</sup> Interestingly, the narwhal and the Beluga whale which diverged ~6 Ma could be differentiated using one of the selected markers, peptide COL1A2T69 ( $\alpha$ 2 (I) residues 801-825).

Overall, the range of modern collagens from different taxa studied to date indicate that the observed protein sequence typically varies by a few amino acid substitutions between different mammalian genera that have diverged ~5-

7 million years ago, with only a small number of exceptions, such as amongst the elephants.

### Fingerprinting collagen in Prehistoric bones

Of 74 ancient fossils sampled (restricted to <1 gram to avoid excessive sample destruction), 72 yielded an acid-insoluble collagen residue. The insoluble residues were analysed by MALDI-TOF mass spectrometry and almost all produced collagen peptides, the oldest was ~1.5 Ma.

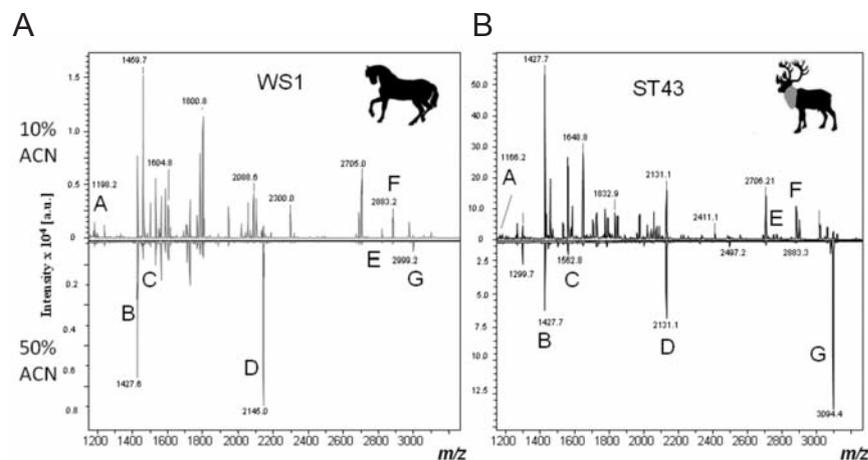
All Holocene samples showed excellent collagen preservation in which even the non-helical telopeptides were detected and amino acid racemisation levels remained low (Asx D/L <0.15; Table 1). In all of the Holocene samples <10 Ka, collagen fingerprints were obtained that allowed for taxonomic identification via the ZooMS approach. In these 25 specimens from six archaeological sites, ZooMS confirmed the identity of the morphologically-determined species in all except one (Table 1). Although it only identifies to genus level, for many taxa, including roe deer (*Capreolus sp.*) and boar (*Sus sp.*), the species could be inferred due to the lack of closely-related members present in Britain at this time. A juvenile caprine skeletal element from Carsington Pasture Cave was identified as sheep (*Ovis sp.*), but aurochs and cattle specimens could only be identified to the sub-family *Bovinae*. The only analysis of a Holocene sample that conflicted with morphological identification was a highly degraded Iron Age bone sample from Wetwang Slack, initially thought to be human with an unusual herbivorous isotopic signature but later identified as horse (Figure 1A).

Sufficient peptide markers were observed in the spectra of most of the Late Pleistocene samples dating back to ~125 Ka (Table 2), allowing the confirmation of canid specimens (i.e., dog or fox) to *Canis sp.* and several samples limited to *bovid* as bovine. A suspected canid from Uphill Cave was identified as red fox (*Vulpes vulpes*) rather than arctic fox (*Alopex lagopus*), which although overlap can indicate subtly different environmental conditions yet are often difficult to separate using morphological criteria alone. Unidentifiable Late Pleistocene fragments from Steetley Wood Quarry were identified as bovine and reindeer (Figure 1B). The analyses of the helical peptides from these samples were all comparable with those of modern bone, but it is within this period when a substantial loss of the non-helical collagen telopeptide is detected (Table 2). Although the amino acid racemization values of all the material dredged from the North Sea were low, the average Asx D/L value for this set of samples was greater than the younger Holocene specimens reflecting the

**Table 1. Morphological and biochemical identifications of Holocene samples compared with the recorded data including number of peaks observed in MALDI-TOF-MS analysis (from tryptic digest only; t = telopeptide method carried out, ts = telopeptide observed), Glycine concentrations and Asx racemization per sample. New biochemical identifications are marked with underlined text, shaded rows highlight errors in morphological identification.**

Sample	Age	Thermal age	No. Peaks	Gly Conc. (nmol/mg)	Asx D/L	PMF identification	Morphological identification
ET2	2500	1584	111ts	217.26	0.14	Bovine	Aurochs
WS2	2200	2115	133	212.92	0.12	Horse	Human
NF1	3370	2778	182	1121.72	0.09	Bovine	Aurochs
CLH1	3600	3255	169	336.89	0.10	Sheep	Sheep
CLH2	3600	3255	158	365.69	0.11	Sheep	Sheep
CLH3	3600	3255	179	888.59	0.13	Bovine	Cattle
CLH4	3600	3255	173	348.32	0.11	Bovine	Cattle
CLH5	3600	3255	156	236.92	0.10	Red/Fallow Deer	Red deer
CLH6	3600	3255	160	204.44	0.11	Canis	Dog
CLH7	3600	3255	158	264.59	0.10	Canis	Dog
CLH8	3600	3255	191	331.71	0.11	Bovine	Cattle?
CLH9	3600	3255	196	249.90	0.11	Bovine	Cattle?
CLH10	3600	3255	147	297.71	0.11	Bovine	Cattle?
CPC1	5936	3761	161ts	941.15	0.09	Bovine	Aurochs
CPC2	5936	3761	130ts	821.65	0.10	Sheep	Caprine
CPC3	5936	3761	192ts	964.10	0.11	Pig	Pig
CPC4	5936	3761	133ts	417.28	0.10	Horse	Horse
CPC5	5936	3761	154ts	692.49	0.13	Roe Deer	Roe Deer
CPC6	5936	3761	169ts	360.56	0.10	Bovine	Cattle
TP1	6174	6101	177ts	213.50	0.12	Bovine	Aurochs
ET1	10000	6932	171t	967.39	0.12	Bear	Bear
ET3	10000	6932	158ts	679.27	0.11	Bear	Bear
ET4	10000	6932	112ts	1044.95	0.10	Bear	Bear
ET5	10000	6932	164t	619.37	0.13	Cat	Cat
ET6	10000	6932	118t	66.31	0.15	Distinct from known taxa (most similar to cat)	Lynx

ET, etches cave; WS, wetwang slack; NF, north ferriby; CLH, cladh hallan; CPC, carsington pasture cave; TP, totty pot cave.



**Figure 1. (A) Example MALDI-MS spectra of the 10% and 50% acetonitrile fraction of a collagen digest from morphologically-unidentified bone fragment WS1 (left) identified as horse, and (B) ST43 (right), identified as reindeer. A = COL1A2T85, B = COL1A2T43, C = COL1A2T45, D = COL1A2T69, E = COL1A2T41/42, F = COL1A1T55/56, and G = COL1A2T67.**

expected increase in protein degradation. It is noteworthy that the ~ 45 Ka bison specimens from Titan gave a good (helical) collagen fingerprint despite having a particularly high Asx D/L value (0.39).

Collagen peptide fingerprints were obtained from all of the Middle and Lower Pleistocene samples analysed (Table 3) including the Cromerian deposits (~500-800 Ka) and the older Weybourne Crag deposits (~1.5 Ma). The exceptional preservation from fine-grained estuarine deposits at West Runton (~650 Ka) and Happisburgh (~900 Ka) is likely related to the extreme compactness of the deposits that offer very low permeability and thus anoxic

nature of the deposit.<sup>17</sup> The morphologically-unidentifiable fragments were identified as deer and bovine (Figure 2) and the suspected large-mammals were identified as elephantid but it was not possible to distinguish between the elephants (*Elephas* and *Loxodonta*) and the mammoths (*Mammuthus*) as discussed elsewhere.<sup>4</sup> The high quality collagen fingerprint obtained from the fish specimen (Supplementary Figure 2) was perhaps most surprising given that they are typically poorly preserved in most assemblages in comparison to more robust mammalian bone fragments. Identification of morphologically-unidentifiable fragments of bone from the archaeological

site of Happisburgh (Site 3) highlight the use of collagen fingerprinting in understanding the environments encountered by some of the earliest humans in Britain (Figure 2).<sup>18</sup>

## Discussion

### How long could collagen survive in fossils?

We have estimated the effective activation energy (Ea) of collagen loss from bone to be 173 kJ mol<sup>-1</sup>. Extrapolation from high tempera-

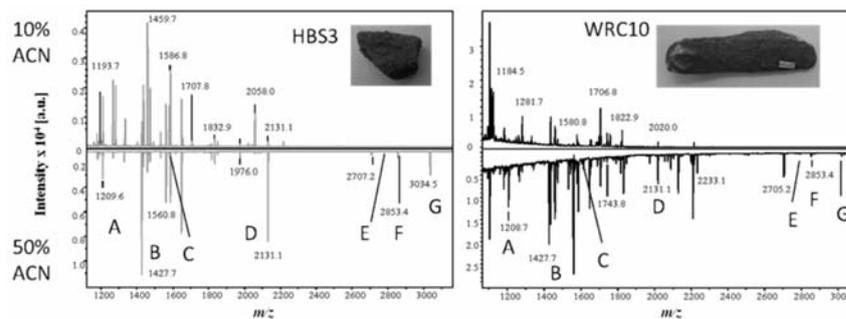
**Table 2. Morphological and biochemical identifications of Late Pleistocene samples compared with the recorded data including number of peaks observed in MALDI-TOF-MS analysis (from tryptic digest only; t = telopeptide method carried out, ts = telopeptide observed), Glycine concentrations and Asx racemization per sample. New biochemical identifications are marked with underlined text; shaded rows highlight errors in morphological identification.**

Sample	Age	Thermal age	No. Peaks	Glycine Conc. (nmol/mg)	Asx D/L	PMF Identification	Morphological identification
FHC1	12000	7342	109	4.18	0.26	Bear	Bear
LOW1	Dredged	9716	156ts	628.83	0.12	Red/Fallow Deer	?
NS1	Dredged	21277	103	405.25	0.15	Distinct from known taxa (most similar to narwhal)	Beluga Whale
NS2	Dredged	21277	25ts	752.88	0.09	Bovine	Bison
NS3	Dredged	21277	168	551.37	0.12	Pig	Wild Boar
NS4	Dredged	21277	141	736.57	0.15	Elephant	Mammoth
NS5	Dredged	21277	115	932.21	0.09	Human	Human
NS6	Dredged	21277	160	616.35	0.12	Distinct from known taxa (most similar to other cetaceans)	Whale
NS7	Dredged	21277	160t	703.85	0.12	Seal*	Harp Seal
NS8	Dredged	9716	167t	311.71	0.11	Bear	Bear
UP1	30000	10937	88t	336.90	0.13	Red Fox	Canid?
UP3	30000	10937	168t	419.84	0.12	Hyaena	Bear
TI1	45800	7186	128t	152.28	0.39	Bovine	Bison
SCC1	79000	7968	83t	699.24	0.10	Reindeer	Reindeer
ST11	66000	9052	118t	587.16	0.11	Canis	Canid?
ST12	66000	9052	179t	608.10	0.11	Canis	Wolf
ST16	66000	9052	186t	474.89	0.14	Reindeer	Reindeer
ST18	66000	9052	89t	187.48	0.16	Bovine	?
ST23	66000	9052	186t	474.89	0.14	Reindeer	Reindeer
ST33	66000	9052	89ts	187.48	0.16	Bovine	?
ST36	66000	9052	128t	187.48	0.17	Bovine	?
ST38	66000	9052	99t	614.62	0.13	Bovine	Bovid
ST43	66000	9052	190t	761.40	0.11	Bear	Artiodactyl?
KC1	125000	19239	125t	369.50	0.13	Rhino	Rhino
KC2	125000	19239	139t	343.50	0.12	Bovine	Rhino
KC3	125000	19239	57t	626.82	0.14	Bovine	Cattle
KC4	125000	19239	169t	618.21	0.12	Bovine	Cattle
KC5	125000	19239	96t	206.41	0.13	Bovine	Bison
KC6	125000	19239	130t	343.46	0.26	Bovine	?
KC7	125000	19239	122t	512.78	0.13	Bovine	?
LAV1	125000	24630	-	19.937	0.54	-	Hippo
LAV2	125000	24630	-	2.283	0.28	-	Hippo

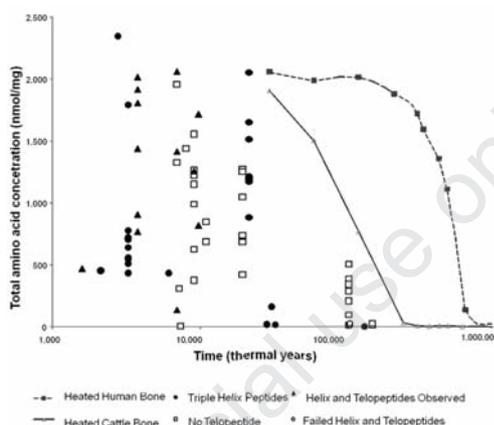
TI, titanshaft; FH, fox hole cave; SCC, stump cross caverns; ST, steetley wood cave; UP, uphill cave; NS, dredged from the North Sea; LAV, lavanham. \*The markers described in [2] are identical for both Harp and Grey seal.

ture experimental decomposition rates using this activation energy suggest that at a constant 10°C (the approximate mean annual air temperature in present-day Britain) it will take between 0.2 and 0.7 Ma years at 10°C for levels of collagen to fall to 1% in an optimal burial environment. The more conservative of these two estimates extends the useful range for collagen sequencing to the Lower Pleistocene (once the long glacial intervals of the British Quaternary are factored into the temperature calculations; Figure 3). Collagen could plausibly be detected at lower concentrations in much older material but likely in a diagenetically-altered state and at levels whereby separation from endogenous and exogenous contaminations is much more time-consuming, costly and perhaps applicable only to atypically large taxa that can offer sufficient fossil material for destructive analysis.<sup>19</sup>

Amino acid yields from bone samples, plotted against thermal age (normalising the *chemical age* of bones by taking into account the thermal history of the site), broadly follow the rate predicted for survival (Figure 3). Estimates of the thermal history of this region were based upon modelled fluctuations in air temperatures<sup>20</sup> scaled using local data from coleopteran assemblages.<sup>21</sup> Extrapolation from high temperature experiments of collagen loss from both mature and immature bone, suggests lower and upper limits for collagen survival at 10°C of 0.2-0.7 Ma. Due to extensive



**Figure 2.** Previously unidentified specimen (WRWC10) from the Weybourne Crag deposits at West Runton identified as bovine (*Bos/Bison*) using collagen fingerprinting (MALDI-MS spectrum from 10% ACN fraction shown). A = COL1A2T85, B = COL1A2T43, C = COL1A2T45, D = COL1A2T69, E = COL1A2T41/42, F = COL1A1T55/56, and G = COL1A2T67.



**Figure 3.** Comparison between predicted and observed loss of collagen scaled to thermal age (collagen years at a constant temperature of 10°C). Experimental data on collagen loss in heated cattle and human bone extrapolated to 10°C reveals considerable variation simply as a function of the age (extent of cross-linking) within the bone, irrespective of any errors in estimation of the rate. Collagen peptides were detected in all but three ancient samples. Half of the remaining samples were tested for the presence of telopeptides which were not detected in the thermally oldest samples. Experimental data from which the collagen survival curves are taken suggests that there is a small residual fraction with a distinct bulk amino acid composition persisting beyond this period.

**Table 3.** Morphological and biochemical identifications of Middle and Lower Pleistocene samples compared with the recorded data including number of peaks observed in MALDI-TOF-MS analysis (from tryptic digest only; t, telopeptide method carried out, ts, telopeptide observed), Glycine concentrations and Asx racemization per sample. New biochemical identifications are marked with underlined text, shaded rows highlight errors in morphological identification.

Sample	Age	Thermal age	No. Peaks	Gly Conc. (nmol/mg)	Asx D/L	PMF Identification	Morphological identification
LAT1	200000	28128	150	7.29	0.15	Elephant	Mammoth
AVE1	200000	30270	143	81.78	0.27	Elephant	Straight tusked elephant
IL1	200000	32067	133	9.08	0.56	Bovine/Musk Ox*	Bovid
KS2	650000	100591	117	49.71	0.35	Bovine	Musk Ox
WR1	650000	100591	167	84.42	0.43	Distinct from known taxa (no markers match)	Fish
WR4	650000	100591	146	46.29	0.23	Elephantid	Steppe mammoth
WR5	650000	100591	146t	194.60	0.17	Deer	Deer
WR6	650000	100591	104	5.79	0.53	Bovine	Large mammal
WR7	650000	100591	111	6.56	0.50	Deer	?
WR8	650000	100591	167t	163.79	0.18	Red/Fallow Deer	?
WR9	650000	100591	177t	244.30	0.16	Red/Fallow Deer	?
WR14	650000	100591	110t	130.40	0.14	Elephantid	Large mammal
HSB3	800000	128,283	86	1.21	0.49	Bovine	?
WRC10	15000000	144862	148	7.84	0.43	Bovine	?
WRC11	15000000	144862	48	6.07	0.51	Bovine	Elephantid?
WRC12	15000000	144862	110	6.37	0.46	Elephantid	Large mammal?
WRC13	15000000	144862	48	9.89	0.61	Bovine	Elephantid?

AVE, aveley; LA, latton; LO, lowestoft; KC, kirkdale cave; IL, ilford; KS, kessingland; WR, west runton forest Bed; HBS, happisburgh site 3; WRC, west runton weybourne crag. \*Due to the poor signal for higher mass peptides in this spectrum; one of the key markers for separating Musk Ox from bovines (Peptide F) was not observed.

and prolonged cold stages in this region, a thermal age of ~0.2 Ma at a constant 10°C is equivalent to a true time-depth that is more than six times this length (~1.5 Ma). It is difficult to extrapolate the upper estimate (~0.7 Ma at 10°C) to an equivalent Quaternary age, due to the lack of pre-Quaternary temperature estimates for Britain but due to the diminishing impact of cold stages, it is estimated that this will be less than the six-fold extrapolation derived from the lower estimate (i.e. <4.0 Ma)

### Collagen survival observed in bone <1.5 million years old

The observation that collagen survives for tens of thousands of years are not surprising given the many studies that have relied upon collagen-derived carbon for accurate dates and dietary information. It is interesting however, that the non-helical collagen telopeptides located at the very ends of each chain and recently considered potentially useful for species identification in archaeological tis-

ues,<sup>8</sup> are lost to the burial environment within a relatively short period of geologic time. These are regions of the protein that do not benefit from as many interchain hydrogen bonds as the helical region and thus will likely be the first to degrade, in a similar manner to other proteins present in bone such as osteocalcin.<sup>10</sup> Although unexpected due to their age (~125 Ka), both samples found from Lavenham failed to show any signals for collagen peptides, one of which is represented by an amino acid composition that differs from that normally seen in fossilised bones containing abundant collagen (Figure 4). Perhaps if enough of the specimen (LAV1) that failed ZooMS analysis but gave an amino acid composition consistent with that of collagen had been sampled it may have been possible to retrieve collagen peptides. Although the poor molecular preservation is likely to be greatly influenced by burial environment (in this case ~125 Ka Pleistocene sands underlying brick-earth), further work needs to be carried out to understand the relative significance of differ-

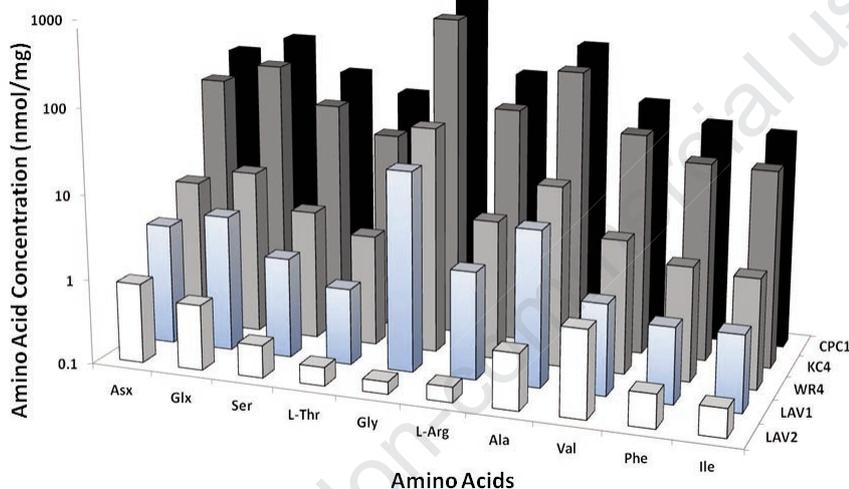
ent factors of a burial environment, not only temperature, but soil type, pH, hydrology and perhaps many more.

There are relatively few reports of collagen isolated from Middle<sup>3,22</sup> and Lower Pleistocene<sup>5</sup> bones but thermal age predictions (Figure 3) would suggest that collagen could survive this long in optimal preservation conditions. Interestingly, some spectra showed greatly diminished signals for peptides of  $m/z$  greater than 2400 (Figure 2), perhaps supporting the idea that collagen degrades through hydrolysis at a first order rate.<sup>9</sup> Alternatively, the signals for these larger peptides could be getting suppressed by a greater abundance of larger humic molecules in the older material.

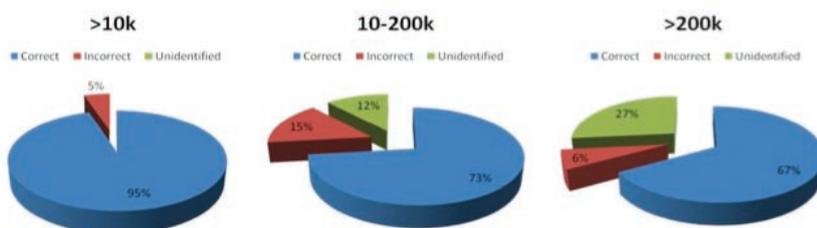
Despite this decrease in signal intensity of some of the larger peptide markers, their persistence in fingerprints from specimens as old as ~1.5 Ma allows for taxonomic identification in most cases. Interestingly, although there is no significant increase in the number of incorrect morphological identifications observed with age, there appears to be much greater amounts of morphologically-unidentifiable material (Figure 5). This indicates that a method of taxonomic identification as robust as this collagen fingerprint approach has great potential to studying specimens from these earlier parts of the Quaternary Period, particularly the Lower and Middle Pleistocene.

When allied to plausible estimates of Pleistocene air temperatures and extrapolation of temperature dependence based upon high temperature laboratory studies, thermal age estimates suggest collagen may plausibly persist at sufficient concentrations for detection in Lower and Middle Pleistocene bone from temperate latitudes. These predictions are placed on a surer footing by direct analysis of collagen peptides from Cromerian Forest Bed and underlying Weybourne Crag samples. However, the failures to obtain collagen from one site (Lavenham) and high success rates from the much older specimens from the Cromerian forest bed stress the importance of local depositional environment on collagen survival.

With only a few exceptions, the collagen peptides enabled genus level identification, which within specific ecological contexts would also permit taxa to be identified to species level. The advantage of collagen fingerprinting over other molecular methods of taxon identification is its potential to be applied to most archaeological and palaeontological assemblages in temperate Quaternary sediments on morphologically indistinguishable fragments. This is particularly useful for older and much less familiar taxa that have less comparative material available and, although this study focuses on the British Pleistocene, is equally applicable worldwide.



**Figure 4.** Bar chart showing amino acid compositions of selected fossil samples. Most specimens show similar compositions indicating the likely dominance of collagen present in at least one of the two samples (LAV1) that failed collagen peptide fingerprint analysis. Sample LAV2 contained detectable concentrations of various amino acids but does not reflect a composition typical of collagen.



**Figure 5.** Summary showing the decreasing confidence in identifications made using morphological criteria over time.

## References

1. Currant A, Jacobi R. A formal mammalian biostratigraphy for the Late Pleistocene of Britain. *Quat Sci Rev* 2001;20:1707-16.
2. Buckley M, Collins M, Thomas-Oates J, Wilson JC. Species identification by analysis of bone collagen using matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry. *Rapid Commun Mass Spectrom* 2009;23:3843-54.
3. Schweitzer MH, Hill CL, Asara JM, et al. Identification of immunoreactive material in mammoth fossils. *J Mol Evol* 2002;55:696-705.
4. Buckley M, Larkin N, Collins MJ. Mammoth and Mastodon collagen sequences; survival and utility. *Geochim Cosmochim Acta* 2011;75:2007-16.
5. Palmqvist P, Grocke D, Arribas A, Fariña R. Paleoeological reconstruction of a lower Pleistocene large mammal community using biogeochemical (13C, 15N, 18O, Sr:Zn) and ecomorphological approaches. *Paleobiology* 2003;29:205-29.
6. Matter P, Davidson FD, Wykoff RWG. The microstructure and composition of some Pliocene fossils. *Comp Biochem Physiol* 1970;35:291-8.
7. Dobberstein RC, Collins MJ, Craig OE, et al. Archaeological collagen: Why worry about collagen diagenesis? *Archaeol Anthropol Sci* 2009;1:31-42.
8. Buckley M, Collins MJ, Thomas-Oates J. A method of isolating the collagen (I) alpha 2 chain carboxyterminal peptide for species identification in bone fragments. *Anal Biochem* 2008;374:325-34.
9. Collins MJ, Riley MS, Child AM, Turner-Walker G. A basic mathematical simulation of the chemical degradation of ancient collagen. *J Arch Sci* 1995;22:175-83.
10. Buckley M, Anderung C, Penkman K, et al. Comparing the survival of osteocalcin and mtDNA in archaeological bone from four European sites. *J Arch Sci* 2008;35:1756-64.
11. Geffen E, Mercure A, Girman DJ, et al. Phylogenetic relationships of the fox like canids: mitochondrial DNA restriction fragment, site and cytochrome b sequence analyses. *J Zoology* 1992;228:27-39.
12. Wozencraft WC, Wilson DE, Reeder DM. *Mammal species of the world: a taxonomic and geographic reference*. Washington, London: Smithsonian Institution Press; 2005.
13. Flynn JJ, Finarelli JA, Zehr S, et al. Molecular phylogeny of the Carnivora (Mammalia): assessing the impact of increased sampling on resolving enigmatic relationships. *Syst Biol* 2005;54:317-37.
14. Lenstra JA, Bradley DG. Systematics and phylogeny of cattle. In: Fries R, Ruvenski, A, editors. *The Genetics of Cattle*. New York: CAB International; 1999. pp. 1-14.
15. Pitra C, Fickel J, Meijaard E, Groves C. Evolution and phylogeny of old world deer. *Mol Phylogenet Evol* 2004;33:880-95.
16. McGowen MR, Spaulding M, Gatesy J. Divergence date estimation and a comprehensive molecular tree of extant cetaceans. *Mol Phylogenet Evol* 2009;53:891-906.
17. Turner-Walker G. The West Runton fossil elephant: A pre-conservation evaluation of its condition, chemistry and burial environment. *The Conservator* 1998;22:26-35.
18. Parfitt SA, Ashton NM, Lewis SG, et al. Early Pleistocene human occupation at the edge of the boreal zone in northwest Europe. *Nature* 2010;466:229-33.
19. Schweitzer MH, Zheng W, Organ CL, et al. Biomolecular characterization and protein sequences of the Campanian hadrosaur *B. canadensis*. *Science* 2009;324:626-31.
20. Bintanja R, van de Wal RS, Oerlemans J. Modelled atmospheric temperatures and global sea levels over the past million years. *Nature* 2005;437:125-8.
21. Coope GR. Quaternary climatic changes-what can we learn from history? *The ESS Bulletin* 2004;2:57-75.
22. Jones A, O'Connell TC, Young E, et al. Biogeochemical data from well preserved 200 ka collagen and skeletal remains. *Earth Planet Sci Lett* 2001;193:143-9.